Clemson University
TigerPrints

All Dissertations

Dissertations

May 2019

Essays in Economic Growth and International Trade

Narendra Raj Regmi
Clemson University, narendra.regmi@gmail.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation
Regmi, Narendra Raj, "Essays in Economic Growth and International Trade" (2019). *All Dissertations*. 2359.
https://tigerprints.clemson.edu/all_dissertations/2359

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

Essays in Economic Growth and International Trade

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Economics

by
Narendra Raj Regmi
May 2019

Accepted by:
Dr. Robert Tamura, Committee Chair
Dr. Scott Baier
Dr. Michal Jerzmanowski
Dr. Gerald P. Dwyer, Jr.

Abstract

The first chapter of this dissertation identifies the decline in mortality risk as an important trigger of demographic transition. To that end, we solve a precautionary demand for fertility model to fit the time series of total fertility rate and average years of schooling in the labor force for the 16 rich countries. Overall, the time series of fertility and average years of schooling for individuals in the labor force generated by our model closely match the actual observations. Furthermore, the out-of-sample prediction of output per worker are also highly correlated with the data. Using the model, we also identify a temporary decline in the price of housing space as the leading cause of baby booms across these countries.

The second chapter employs machine learning techniques to capture heterogeneity in free trade agreements. The tools of machine learning allow us to quantify several features of trade agreements, including volume, comprehensiveness, and legal enforceability. Combining machine learning results with gravity analysis of trade, we find that more comprehensive agreements result in larger trade creation effects. In addition, we identify the specific trade policy provisions that tend to have the substantial effect in creating trade flows. In particular, legally binding provisions on anti-dumping, capital mobility, competition, customs harmonization, dispute settlement mechanism, e-commerce, environmental standards, export restrictions, freedom of transit, import restrictions, institutional arrangements, intellectual property rights, investment, labor standards, public procurement, sanitary and phytosanitary measures, services, subsidies and countervailing measures, technical barriers to trade, telecommunications, and transparency tend to have the largest trade creation effects.

Dedication

I dedicate this dissertation to my grandfather, Late Upendra Raj Regmi, and uncle, Late Rajendra Raj Regmi, whose life principles always motivated me to be a go-getter and pursue a terminal degree in a field that I love the most. I also dedicate this work to my grandmother, Bishnu Maya Regmi, parents, Yogendra Raj Regmi and Chhetra Kumari Regmi, and my brother Gyanendra Raj Regmi.

Acknowledgments

This dissertation has been greatly improved by the assistance of numerous people. I would like to thank my advisor, Dr. Robert Tamura, for his scholarly guidance in the field of fertility, human capital, and economic growth. I am also grateful to Dr. Scott Baier for his guidance in helping me understand the gravity model of trade despite his hectic schedule as the Chair of the Economics Department. My committee members Dr. Gerald P. Dwyer, Jr. and Dr. Michal Jerzmanowski also deserve a big gratitude for their helpful comments and suggestions.

Next, I would like to thank my undergraduate professor, Dr. Jerome Savitsky, as well as my good friend, Dr. Sanjeev Khatiwada, who encouraged me to pursue a doctorate degree in economics.

I would also like to remember my uncle, Bishnu Prasad Sigdel, and aunt, Bhagirathi Sigdel, for their continuous support.

Finally, I would like to thank my friends at Clemson for their support and encouragement throughout these years. In particular, I would like thank Aastha, Steven, Roxana, Ben, Liuna, Mahesh, and Paresh for their support during all these years.

Table of Contents

List of Tables

List of Figures

Chapter 1

Fertility, Human Capital, and Economic Growth: International Evidence (with Dr. Robert Tamura)

1.1 Introduction

During the last two centuries, most parts of the world experienced a dramatic decline in fertility for the first time, a phenomenon collectively known as demographic transition. Beginning with wealthier nations in the late nineteenth century, the middle-income and low-income countries began to experience similar dramatic declines in fertility and mortality in the middle and late twentieth century, respectively. For example, on average, women from Western countries had a fertility rate of 2.5 in 1928. By 1960, Southern Europe and Central and Eastern Europe reached that same rate. By the early twenty-first century, most countries in every region of the world except for North Africa and Sub-Saharan Africa reached the fertility rate of 2.5. During the same period, the majority of countries also experienced a sustained growth in output per capita for the first time in history. Therefore, understanding the triggers of demographic transition and the role of demographic transition in generating economic transition is an important question in the field of economic growth. This paper provides empirical support for the role of declining mortality risk in generating demographic transition across 16 rich countries.

This paper makes two important contributions. First, this paper presents a mechanism through which the decline in mortality risk leads to a fall in fertility and a corresponding rise in human capital accumulation. When mortality risk is high, the uncertainty of child survival gives rise to "hoarding" of children motive for parents. As mortality risk declines, parents reduce their precautionary demand for children, which in turn lowers the price for child quality. This allows parents to invest more resources per child leading to an accelerated accumulation of human capital. Using this particular mechanism, we explain the secular decline in fertility as well as a secular rise in human capital observed across the rich countries in our sample.

Second, as in Tamura and Simon (2017) we attribute baby booms to a fall in the price of housing space. Similarly, we identify fall in the cost of schooling to explain the ostensibly paradoxical rise in the years of schooling observed during the baby boom years. This phenomenon runs counter to the quality-quantity trade-off highlighted by Becker and Lewis (1973) because as the quantity of children increases, the price of child quality rises, which in turn would imply a reduction in the average years of schooling. By calibrating an endogenous model of fertility and human capital to fit the observed series of fertility and average years of schooling for 16 rich countries, we identify a temporary decline in the price of housing space, along with a fall in the cost of schooling, as a possible explanation for the phenomenon.

Murphy, Simon, and Tamura (2008) use this same precautionary model of fertility to fit the differential baby booms and the rising level of schooling attainment across US states by allowing the price of housing space and the cost of schooling to vary over time. The time-varying price series of housing space chosen to fit the data on fertility and years of schooling closely fits the average population densities of the US states during the same period. Similarly, Tamura and Simon (2017) use the same precautionary model to fit the time series of fertility and years of schooling for 21 OECD countries for the last two centuries. They also find strong empirical support for the time-series of calibrated cost of schooling when they use historical data on education expenditures as a share of GDP to proxy for the cost of schooling in these countries.[1] Overall, both our model-generated

[1]Tamura, Simon, and Murphy (2016) use a similar model of precautionary fertility to explain black and white fertility and years of schooling for 1800 (1820 for black) to 2000 at the U.S. state level.

series of fertility and our model-generated series of average years of schooling closely match the corresponding data, and our out-of-sample predictions of real output per worker are also highly correlated with the data.

In Section 1.2, we present literature review on studies on fertility, human capital, and baby booms. Section 1.3 presents the model, and the calibration procedure. Section 1.4 explains data sources. Section 1.5 presents a brief discussion of the data on fertility, mortality and years of schooling used to calibrate the model. Section 1.6 presents our calibration results and goodness of fits results, and Section 1.7 concludes.

1.2 Literature Review

1.2.1 Existing Theories of Fertility and Economic Growth

Prior to Becker (1960), economists considered fertility to be outside the scope of economics. This was mainly because earlier studies that did not control for the differential knowledge of contraceptives among parents had found either a negative or no relationship between family income and fertility (Becker (1960)). By modeling children as consumer durables and by introducing child quality into parent's utility function, Becker (1960) shows that the demand for children (quantity) and child quality rise with income.

Becker and Barro (1988) link fertility choice to the theory of economic growth. They analyze fertility within an intergenerational model in which altruistic parents derive utility from the number of children as well as the utility of children. The theory suggests that fertility should be positively related to interest rates. They use this framework to account for the fall and rise of fertility during the Great Depression and the post-war baby boom in the United States.

Becker, Murphy, and Tamura (1990) link the dynastic Becker and Barro (1988) fertility model with human capital. By assuming that parents' discount rate on per capita consumption of future generations is negatively related to fertility and that the rate of return on investment in human capital is increasing in its stock, Becker, Murphy, and Tamura (1990) show that there exists two distinct development regimes. The first is a "Malthusian" regime characterized by low levels of human capital investment and high levels of fertility. The second regime is a "Development" regime where parents choose low levels of fertility and make high levels of human capital investment. By putting human capital at its core, the model was successful in jointly explaining low fertility and high rates of economic growth, typical of modern economies.

Becker, Murphy, and Tamura (1990) motivated a stream of literature that examine the triggers of transition from “Malthusian” regimes to “Development” regimes. Galor and Weil (1998) present a “unified” endogenous growth model in which the economy endogenously evolves to a Development regime from a Malthusian regime. In the first stage, Malthusian phase, technological progress is very slow, and advancements in technology leads to a proportional increases in output and population. Due to the positive relationship between population and technological progress in the Malthusian phase, technological progress gradually rises and leads to an increase in the rate of return on human capital. The rise in return to human capital encourages parents to increase investment in human capital of their children, which in turn accelerates technological progress and generates a virtuous circle. This cycle ultimately induces demographic transition and a state of sustained economic growth.

Galor and Moav (2002) present an interesting hypothesis that the evolutionary pressures played a vital role in the transition from stagnation to growth. They argue that natural selection gradually reshaped the composition of the population in a way in which off-springs carryover the traits that are conducive to higher levels of income. In essence, natural selection caused an increase in the representation of traits such as preferences for child quality over quantity that ultimately triggered the demographic transition.

1.2.2 Literature Review on Baby Booms

One of the interesting fertility dynamics in the last century is the occurrences of baby booms in many parts of the world. There are several explanations put forward for baby booms and baby busts in the literature. Greenwood, Seshadri, and Vandenbroucke (2005) identify the invention of labor-saving household capital, such as modern appliances that decreased the cost of fertility during the middle of the century, caused the baby boom in the United States and other Western countries. Doepke, Hazan, and Maoz (2015) show that the increased demand for female labor in World War II was the main reason for baby boom in the United States. The female workers, who replaced male workers in the labor market while the male workers were fighting the war, continued to work even long after the war. When the younger women entered the labor market, they faced increased competition from the older and more experienced female workers. With bleak prospects in the labor market, the younger female cohorts opted to marry earlier and bore more children instead.

These existing explanations for the cause of baby booms experienced by the United States and other Western countries may not be applicable for all countries in the world. Countries that did not actively participate in the Second World War such as India, Jamaica, Sri Lanka, Myanmar, and Cambodia also experienced baby booms and rising educational achievements during their baby boom years. In addition, the argument that the invention of labor-saving appliances decreased the cost of fertility does not explain baby booms in countries that had very little access to electricity. According to the World Bank, only 43% of Indians had access to electricity in 1990s, which is the earliest electrification data publicly available for India World Bank Group. Electrification rates would have been much lower during the 1940s and 1950s when India experienced its baby boom.

Albanesi and Olivetti (2014) is the only paper that presents a model capable of explaining baby booms and concurrent rising educational achievements due to the fall in maternal mortality rate that may be applicable to many countries in the world. They show that the differential declines in maternal mortality across countries, as well as the differential timing of the decline, are consistent with the differential magnitudes and timing of international baby booms.

In this paper, we present a model that jointly explains these fertility dynamics, human capital accumulation and real output per worker in 16 countries. As in Tamura (2006), Murphy, Simon, and Tamura (2008) and Tamura and Simon (2017), we identify secularly declining young-adult mortality as the primary source of demographic transition. However, unlike the existing literature on baby booms, we identify a falling price for housing space as the leading cause of the baby boom and a falling cost of schooling as the source of rise in schooling during the baby boom years. Overall, the model can fit the data on fertility and average years of schooling in the labor force, and also serves to identify the time series of the price of housing space and the cost of schooling. In addition, our model generates data on real output per worker which are highly correlated to actual observations on real output per worker in the 16 countries.

1.3 Model

1.3.1 Parental Preferences

A person lives up to 2 periods and each period is 40 years in length. Parents make decisions in the face of infant mortality and young mortality, death before age 35. A child who survives the first period is assumed to live completely through the second period. Parents choose gross fertility,

x_t, a composite consumption good, c_t, the amount of housing space for each child, S_t, and the "quality" of each child in terms of the level of human capital investment, τ_t, conditional on parental human capital stock, h_t, the probability of young-adult mortality, δ_t, the price of living space r_t, and the cost of schooling κ_t.

Parental preferences are given by

$$\alpha \left(c_t^{\psi} S_t^{1-\psi} \right)^{\varphi} \left[(1-\delta_t)x_t - a \right]^{1-\varphi} + \Lambda h_{t+1}^{\varphi} - \frac{\beta_t \delta_t^{\nu_t}}{\left[(1-\delta_t)x_t - a \right] (1-\delta_t)^{\varepsilon}}, \tag{1.1}$$

where ψ represents the expenditure share between consumption and housing space. We assume that parents only care about net fertility, $(1-\delta_t)x_t - a$, and in order to ensure a lower bound on fertility, we impose $a \geq 0$. The last term, where $\epsilon > 0$, captures the precautionary demand for fertility, similar to Kalemli-Ozcan (2002, 2003), and Tamura (2006), and the pair (ν_t,β_t) represents preferences on the precautionary demand for children. During times of high mortality, parents have more children than necessary to produce the desired number of survivors. [2] As young-adult mortality rate declines, parents' precautionary demand for children falls, and in the limit where mortality goes to zero, the precautionary demand goes to zero as well.

1.3.2 The Parental Budget Constraint

The budget constraint facing a typical parent is given by

$$c_t + r_t x_t S_t = h_t \left[1 - x_t \left(\theta + \kappa_t \tau_t \right) \right], \tag{1.2}$$

where θ is the fixed time cost of raising each child, τ_t is the time spent educating children, κ_t is the efficiency of education time, and r_t is the price per unit of space. Parents divide their time between the labor market and raising children. We also assume that parents choose the size of housing, S_t, and gross fertility, x_t, at the same time.

1.3.3 Human Capital Accumulation Technology

The human capital of the next generation is related to parental human capital, h_t, human capital investment, τ_t, and the frontier human capital in the world, $\overline{h}_t$. The evolution of human

[2] These preferences were used in Tamura and Simon (2017) and are similar to those in Tamura et al. (2016). When young-adult mortality is zero, the preferences in this paper are identical to those in Tamura et al. (2016).

capital is thus given by

$$h_{t+1} = A\overline{h}_t^{\rho_t} h_t^{1-\rho_t} \tau_t^{\mu}. \tag{1.3}$$

The functional form of the international spillover parameter, ρ_t, is given by

$$\rho_t = \min\{.5, .5(\frac{\overline{\tau}_t}{.38125}).\}, \tag{1.4}$$

where $\overline{\tau}_t$ is the average parental time spent in educating each child. As long as the years of schooling is positive and the spillover parameter is positive ($\rho_t > 0$) , the children can access and benefit from the state-of-the-art human capital in the world. The international spillover is maximized at $\overline{\tau} = .38125$ or at 15.25 years of schooling for a 40 year period. This human capital accumulation technology is similar to Tamura (2006) and Tamura, Simon, and Murphy (2016). We are assuming U.S. to be the frontier human capital country, so that $\overline{h}_t$ for all countries is equal to the human capital level of the US in year t.

1.3.4 Model Solution

We substitute (1.3) and (1.2) into (1.1) and differentiate to produce the three Euler equations determining optimal choices of human capital investment, fertility and housing space given by

$$\begin{aligned}
\frac{\partial}{\partial \tau} &= \psi\varphi\alpha c_t^{\psi\varphi-1} S_t^{(1-\psi)\varphi} \left[(1-\delta_t)\, x_t - a\right]^{1-\varphi} h_t x_t \kappa_t = \Lambda\mu\varphi A^{\varphi} (\overline{h}_t^{\rho} h_t^{1-\rho})^{\varphi} \tau_t^{\mu\varphi-1}, & (1.5) \\
\frac{\partial}{\partial x} &= \psi\varphi\alpha c_t^{\psi\varphi-1} S_t^{(1-\psi)\varphi} \left[(1-\delta_t)\, x_t - a\right]^{1-\varphi} \left[h_t \left[\theta + \kappa_t \tau_t\right] + r_t S_t\right] \\
&= (1-\varphi)\, \alpha c_t^{\psi\varphi} S_t^{(1-\psi)\varphi} \left[(1-\delta_t)\, x_t - a\right]^{-\varphi} (1-\delta_t) + \frac{\beta\delta_t^{\nu_t}}{\left[(1-\delta_t)x_t - a\right]^2 (1-\delta_t)^{\varepsilon-1}} & (1.6)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial S} &= \psi\varphi\alpha c_t^{\psi\varphi-1} S_t^{(1-\psi)\varphi} \left[(1-\delta_t)\, x_t - a\right]^{1-\varphi} r_t x_t \\
&= \alpha (1-\psi)\, \varphi c_t^{\psi\varphi} S_t^{(1-\psi)\varphi-1} \left[(1-\delta_t)\, x_t - a\right]^{1-\varphi}. & (1.7)
\end{aligned}$$

We can then solve for c_t as a function of S_t and x_t. Therefore,

$$c_t = \left(\frac{\psi}{1-\psi}\right) r_t x_t S_t. \tag{1.8}$$

Substituting this relationship into the budget constraint (1.2) produces

$$r_t x_t S_t = (1-\psi)\, h_t \left[1 - x_t \left(\theta + \kappa_t \tau_t\right)\right]. \tag{1.9}$$

Substituting this back into the objective function yields

$$\max_{x_t,\tau_t} \left\{ \begin{array}{c} \alpha\,(\psi)^{\psi\varphi} \left(\frac{1-\psi}{r_t x_t}\right)^{(1-\psi)\varphi} \left(h_t \left[1 - x_t \left(\theta + \kappa_t \tau_t\right)\right]\right)^{\varphi} \left[(1-\delta_t)\, x_t - a\right]^{1-\varphi} \\ +\Lambda h_{t+1}^{\varphi} - \frac{\beta \delta_t^{\nu_t}}{(1-\delta_t)^{\varepsilon}} \left[(1-\delta_t) x_t - a\right]^{-1} \end{array} \right\}. \tag{1.10}$$

The budget constraint is non-convex in this setup because fertility interacts with both housing space and human capital. Hence, we cannot obtain analytical comparative statics results. Instead, we employ numerical solution methods to solve the problem. As in Tamura (2006), Tamura, Simon, and Murphy (2016), and Tamura and Simon (2017), we take advantage of the fact that for given fertility, x, the problem is globally concave in (c, S, τ). Therefore, we use a grid over possible values of fertility, ranging from a minimum of $\frac{a}{1-\delta_t}$ to the biological maximum of $\frac{1}{\theta}$, solve for the optimal human capital investment for each level of fertility, and choose the level of fertility that yields the highest utility.

We solved the model annually and produced fertility and human capital investment for that birth cohort. Section 1.4 describes the procedure to compute human capital and years of schooling in the labor force. We allow the taste parameter pair (β_t, ν_t), rental rate (r_t), and schooling efficiency parameter (κ_t), to vary with time for each country in order to fit total fertility rate (x_{it}) and average years of schooling in the labor force in time t.

1.3.5 Stationary Values

In this section we examine the stationary solution. We assume that the stationary fertility rate is 1. Examining the Euler equation with respect to fertility when mortality risk is 0, the parameter restriction on a as a function of the stationary human capital investment rate, $\overline{\tau}$, and other parameters of the model is given by

$$a = 1 - \frac{(1-\varphi)(1-[\theta+\overline{\tau}])}{\varphi(1-\psi(1-[\theta+\overline{\tau}]))}. \tag{1.11}$$

Similarly, the implicit function determining the stationary human capital investment rate, $\overline{\tau}$, is given by

$$1 = \frac{\Lambda\mu\left[Ar^{1-\psi}\right]^{\varphi}(1-\theta-\overline{\tau})^{1-\varphi}}{\alpha\left[\psi^{\psi}(1-\psi)^{1-\psi}\right]^{\varphi}(1-a)^{1-\varphi}\,\overline{\tau}^{1-\mu\varphi}}. \tag{1.12}$$

Under the balanced growth path, the frontier human capital, $\overline{h}_t$, is equal to h_t, and the right hand side of (1.12) is constant. With these parameter restrictions and long run mortality of 0, the long run fertility rate, x, will be 1, and the long run human capital investment rate will be $\overline{\tau}$.

Table 1.1 contains the calibrated parameters in the model. The choices are standard. For example, our choice of fixed time cost of rearing a child, $\theta = 0.125$, is consistent with a biological maximum fertility of 8 in an asexual model. So, for a 40 year period a child is 5 years old when he or she enters school. Our choice of $\overline{\tau} = 0.38125$ implies a steady state value of 15.25 years of schooling. The average years of schooling in our sample countries in 2010 is 14.0 years, and the maximum years of schooling is 15.2 years of schooling. Similarly, our choice of r is the average white population density of U.S. states in the year 2000, where the states are weighted by their 2000 white populations as in Murphy, Simon, and Tamura (2008). Similarly, the choice of (A, μ) is consistent with an annualized growth rate of about 1.80% along the balance growth path. The choice of parameters (θ, ψ) along with the long run values of fertility, x, of 1 and human capital investment, $\overline{\tau}$, of 0.38125 (15.25 years of schooling), and the assumed stationary value of the cost of schooling κ of 1, imply a stationary expenditure share of housing of about 19%. The housing expenditure share reported in the *OECD Better Life Index* is 19.3% for the US in 2016, and 21.3% for rest of the 15 countries in our sample. Thus, for both the US and for the 15 other rich countries, the model parameters calibration match the data very well.

1.3.6 Calibration Procedure

In this section, we briefly discuss our calibration procedure. For each country i, we allow the preference pair (β_t, ν_t), rental price, r_t, and educational efficiency, κ_t , to vary in year t in order to jointly fit the data on total fertility rate, x_{it} and average years of schooling in the labor force.

We follow a two step procedure to fit fertility and schooling. First, we allow the price of space r_t and β_t to vary to match fertility. Since the preference pair (ν_t and β_t) is not separately identified, we impose a constant ν_t value of .4 for the years 1950-2000 for all countries. Second, we search for κ_t to match years of schooling. There was some feedback from changing κ_ts on fertility. In the case of enough feedback, we return to search for the preference pair (β_t, ν_t) to match fertility for the κ that matched the years of schooling in the labor force. Then, we return to adjust our κ to match schooling. This iterative procedure converged to price of housing space, r_t, preference pair,(ν_t,β_t), and cost of schooling, κ_t, that fit both schooling and fertility in a given year t.

Figures 1.1 to 1.16 present the actual observations and model solutions for both models for children ever born, years of schooling in the labor force, schooling of the youngest cohort and output per worker for the 16 rich countries. The countries included in this study are Austria, Belgium, Canada, Denmark, France, Finland, Germany, Ireland, Italy, Japan, Newzealand, Norway, Sweden, Switzerland, United States, and United Kingdom.

1.4 Data

The data on output per worker, young schooling, and average schooling in the labor force come from Tamura et al. (2019). The data on mortality come from Tamura (2006). The data on fertility come from Tamura and Simon (2017).

1.4.1 Young-Adult Mortality Risk δ

We specify young-adult mortality δ as the sum of probability of dying between the ages of 1 and 35, $p_{1,35}$ plus one-third of infant mortality rate, $\frac{m}{3}$. We weigh down the effect of infant mortality by a third because an infant death is less costly to replace than a child death, say at age 14. The reasons are twofold. First, parents would have already made significant human capital investment for 14 years. Second, a large proportion of mother's child bearing age would have disappeared by the death of the child at age 14. It is hard to replace the child for the parents.

We calculate the probability of dying between ages of 1 and 35 using Kaplan Maier survival estimation technique, and it is given by

$$p_{1,35} = 1 - \prod_i [1 - P(i)], \tag{1.13}$$

where 1-P(i) denotes the Kaplan-Maier estimation for survival probability at age class i, and i denotes the following age classes: 1-5, 5-9, 10-14, 15-19, 20-24, 25-29, 30-35. For years prior to 1950 for which data were not available, we extrapolate infant mortality rate and probability of dying between the ages of 1 and 35 by regressing the log of the probability on year and year squared. Thus we set

$$\delta = p_{1,35} + \frac{m}{3}. \tag{1.14}$$

1.5 Fitting Procedure

For total fertility rate and schooling of cohort t, we compare the model solutions in year t with actual observation in year t. For years of schooling in the labor force and real output per worker, we solve the model annually for each country and aggregate the model solutions for 40 years to arrive at the measures of these variables in the labor force.

What follows is the discussion of the methodology to arrive at years of schooling in the labor force and human capital in the labor force from model solutions of human capital and years of schooling for birth cohort born in year t. Since each period is 40 years in length, if a child enter school at year t, the child has already been reared for $40\theta = 5$ years. The child enters the labor force 25 years after birth and works for 40 years before retiring from the labor force. Human capital is assumed to be constant over the entire working period. The labor force participation rate is constant over the work life as well, but it can depend on the level of schooling. So, in order to calculate human capital and years of schooling in the labor force, we weight the cohorts by their relative population size and relative labor force participation rate, l given by

$$N_{t+20}l_t = x_t(1-\delta_t)N_t l_t. \tag{1.15}$$

For labor force participation rates, we use the same measure as Turner et al. 2007. For workers with more than secondary schooling, $\tau_t \geq .3$, we assume labor force participation rate, l_t, of 0.91. For workers with less than full primary schooling, $\tau_t < .2$, we assume labor force participation rate of 0.60. Finally, for workers with exposure to secondary school, but less than high school graduate, $.2 \leq \tau_t < .3$, we assume labor force participation rate of 0.82. The average years of schooling thus obtained is compared in year t with actual observations in year t available from Tamura et al. (2019).

1.6 Results and Discussion

1.6.1 Goodness of Fit and Growth Rates of Output per Worker

In this section, we formally evaluate the model solutions with actual observations. Figures 1.1 to 1.16 contain the data and model solutions for children ever born, years of schooling in the labor force, output per worker, and schooling of the youngest cohort. The plots are arranged alphabetically by country. In order to evaluate the performance of our model, we run the regression

$$y_t = \alpha + \beta x_t \tag{1.16}$$

where y_t is year t observation on either children ever born or average years of schooling in the labor force or schooling of the youngest cohort or $ln(\textit{real output per worker})$; x_t is the year t model children ever born, or the model average years of schooling in the labor force between the ages of 20 to 65, or model schooling of the youngest cohort, or model $ln(\textit{output per worker})$. Under the null hypothesis that the model fits the data perfectly, α equals 0 and β equals 1.

Table 1.2 contains the regression results. The row marked with p in each table denotes the p value of the joint hypotheses. In each of these regressions, we allow for heteroskedasticity and panel auto-correlation. The model does an excellent job of fitting total fertility and the average years of schooling. The first two columns of Table 1.2 presents the regression results on fertility and average years of schooling in the labor force. The coefficient on fertility is 0.91 and the coefficient on average years of schooling is 0.97. The constant terms are 0.31 and 0.27 respectively. The closeness of the coefficients to 1 and the constant terms to 0 implies that the model fits the data very well. The third column reports the results for schooling of the youngest cohort, the coefficient for which is also quite high, about 0.73. The fourth column reports the results for log income per worker, and the slope coefficient is very close to 1, about 0.97, with constant term also very close to 0.

Table 1.3 reports the regression results of growth rates of income per worker against the model income per worker. The coefficient is quite high, 0.80, and highly significant. It is also worth mentioning that nowhere in our calibration exercise do we choose our parameters to directly fit the data on income per worker. So, the fact that not only is the level but also the growth rates of log of income per worker highly correlated with the data counterpart implies that the mechanism we identify in this paper is useful to explain demographic transition and the subsequent economic growth.

1.6.2 Baby Booms

Another important result of the calibration exercise in this chapter is that we identify a fall in the price of housing space as the leading cause of baby boom. In addition, we identify a concurrent fall in the cost of schooling to explain the rise in years of schooling observed during baby boom years. Appendix A reports the time series of rental price, schooling efficiency, and the taste parameter, ν_t and β_t, consistent with the time series of total fertility rate and average years of schooling in the labor force. All countries that experienced baby booms in our sample also saw a decline in the price of housing space during the decade of the baby boom. It is also the case that there was a reduction in the cost of schooling, κ_t during the baby boom years for the countries experiencing baby booms.

1.7 Conclusion and Future Work

As in Tamura (2006) and Tamura and Simon (2017), this paper provides provide further empirical support to the decline in mortality risk as an important trigger for demographic transition. In particular, we find that the decline in mortality can explain the decline in fertility and a secular rise in human capital investment observed in the developed countries. What parents care about is expected net fertility, and as mortality falls, gross fertility falls. The fall in gross fertility allows parents to spend more resources per child thereby raising the child's human capital. The model can not only fit the time series of fertility and years of schooling across the 16 rich countries extremely well, the out-of-sample model prediction of output per worker for these countries are also highly correlated with the data. Using the model, we also identify a decline in the price of housing space as the principal cause of baby boom. The concurrent fall in the cost of schooling is also identified in this paper to explain the secular rise in the average years of schooling in the labor force. Extension of this work will include calibrating this model to middle-income and low-income countries.

Table 1.1: Parameter Values & Calibration

parameter	value	parameter	value	parameter	value
α	0.275	μ	0.085	A	1.55
ψ	0.660	$\overline{\tau}$	0.38125	p	1.000
φ	0.550	a	0.40073833	r	1.529679
θ	0.125	Λ	2.014672872		

Calibration

variable	model	min	max	avg	notes
fertility	2.00	1.4	2.1	1.8	*HDR 2014*
schooling	15.25	13.0	15.2	14.0	Tamura et al. (2019)
annual growth rate	1.80%			1.80%	US 1840-2000

[1] *HDR 2014*: *Human Development Report 2014*.

[2] The growth rate is given by $\ln(A\overline{\tau}^{\mu})/20$. Annual growth of real output per worker from 1840-2000, Turner, Tamura, and Mulholland (2013).

[2] *OECD Better Life Index*.

[3] *Penn World Tables*.

Table 1.2: Pooled Regression Results of Actual Observations on Model Solutions

	Total fertility Rate	Average Years of schooling	Young Schooling	ln(Income)
β	0.9144***	0.9645***	0.7287***	0.9710***
	(0.0296)	(0.0129)	(0.0302)	(0.0245)
α	0.3089***	0.2665***	1.5739***	0.2102***
	(0.0903)	(0.0798)	(0.243)	(0.2286)
$\bar{R}^2$	0.9322	0.9836	0.8006	0.9211
N	479	369	369	339
p	0.0038	0.0045	0.0000	0.3724

Table reports pooled regressions with clustered errors on the country. *** 1%, ** 5%, * 10%. The row, marked p, is the p-value on the null hypothesis that $\beta = 1$ and $\alpha = 0$.

Table 1.3: Regression of Growth Rates

	gy
gh	0.8007***
	(0.0943)
α	0.0024***
	(0.0151)
$\bar{R}^2$	0.079
N	323

Table reports fixed effects regression of growth rates of output on growth rates of model human capital.*** 1%, ** 5%, * 10%.

1.8 Model Series and Actual Series

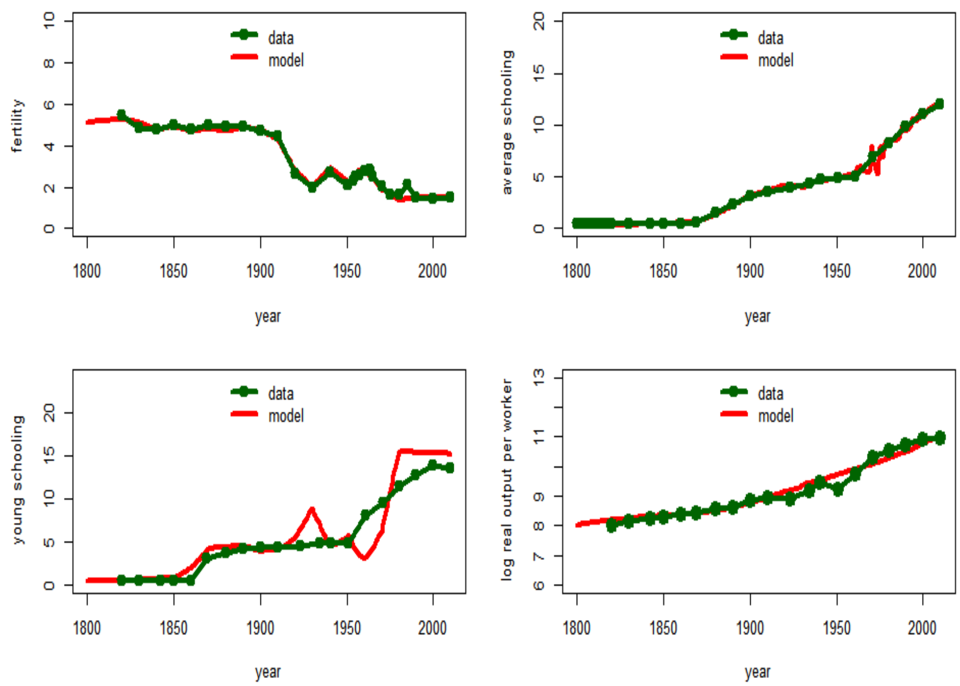

Figure 1.1: Fertility, schooling, real output per worker, young schooling (clockwise): Austria

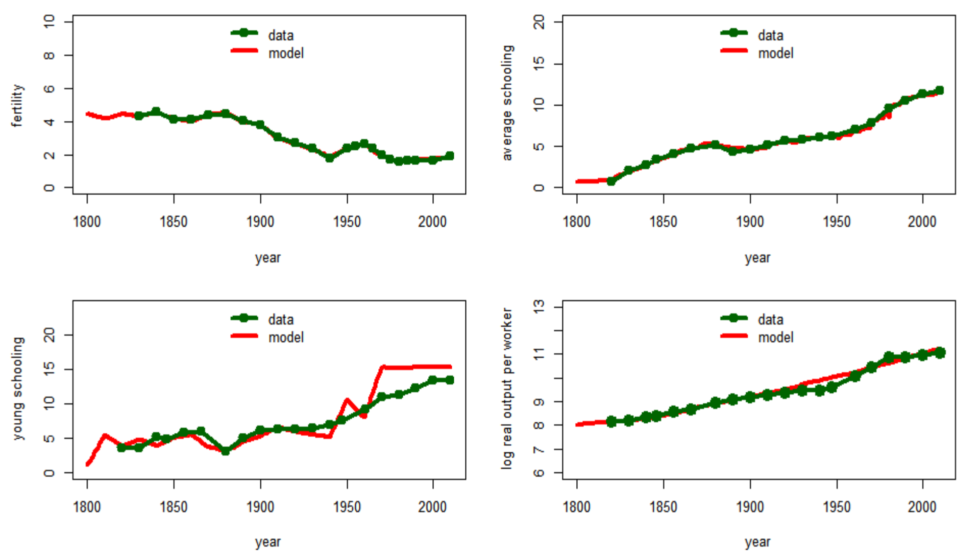

Figure 1.2: Fertility, schooling, real output per worker, young schooling (clockwise): Belgium.

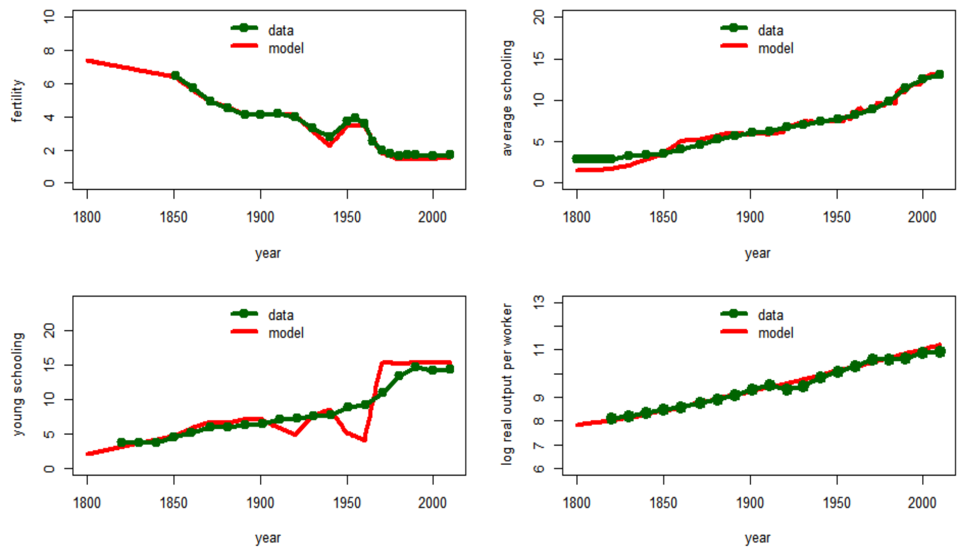

Figure 1.3: Fertility, schooling, real output per worker, young schooling (clockwise): Canada.

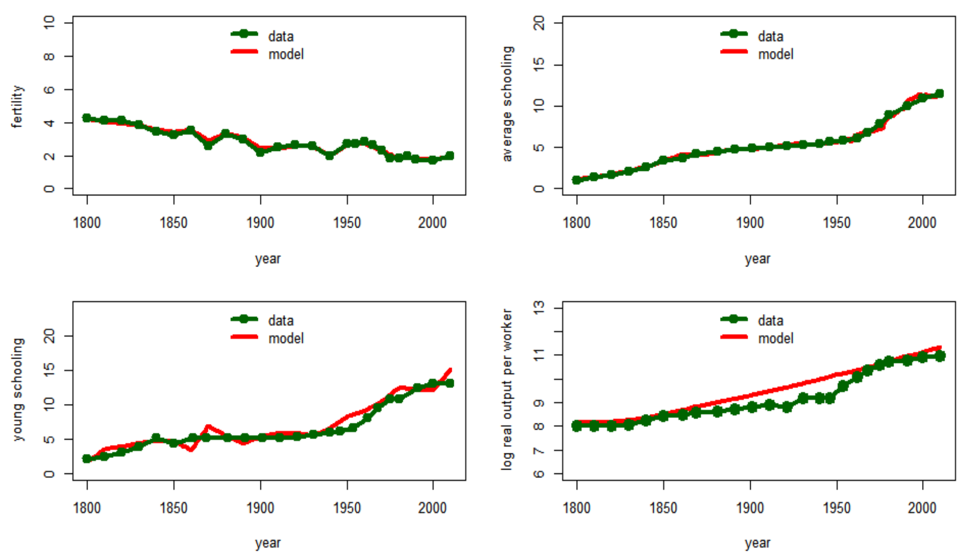

Figure 1.4: Fertility, schooling, real output per worker, young schooling (clockwise): Denmark.

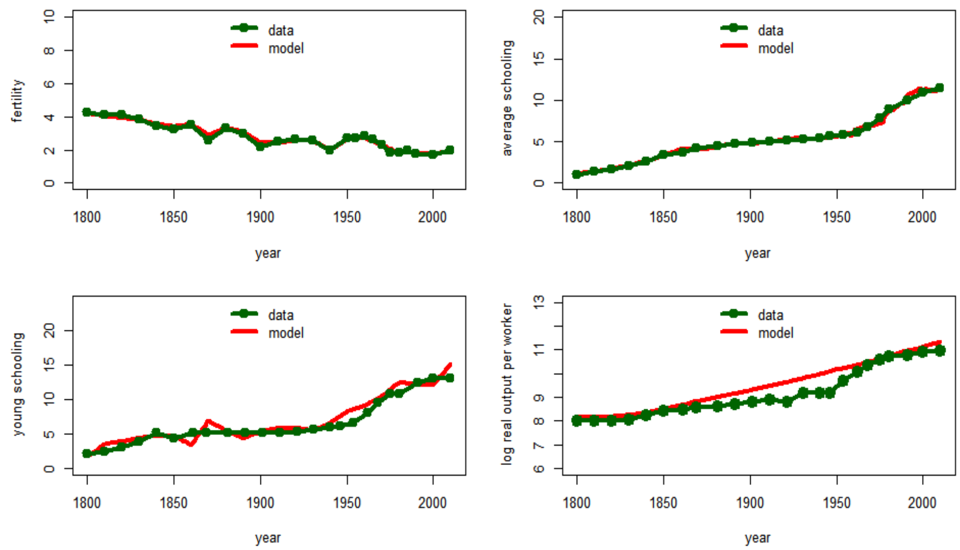

Figure 1.5: Fertility, schooling, real output per worker, young schooling (clockwise): France.

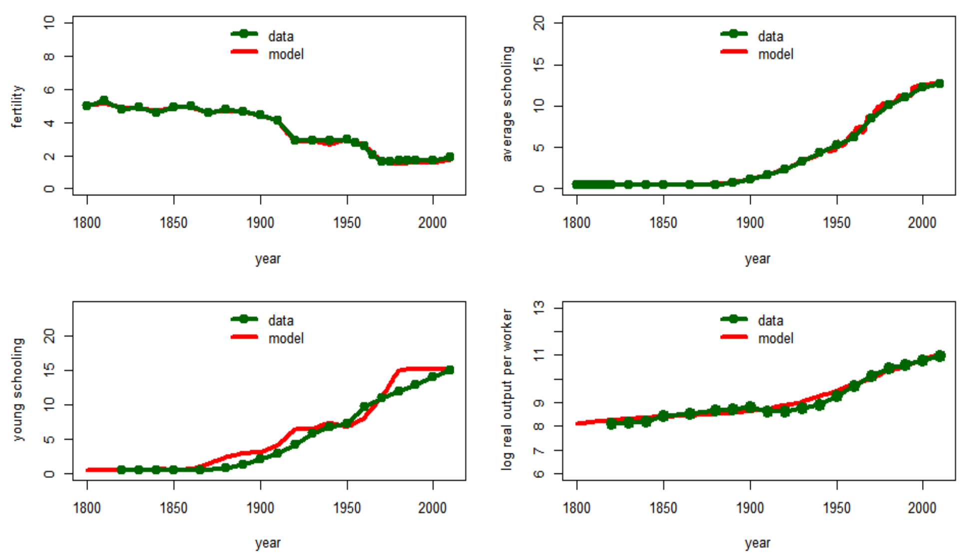

Figure 1.6: Fertility, schooling, real output per worker, young schooling (clockwise): Finland.

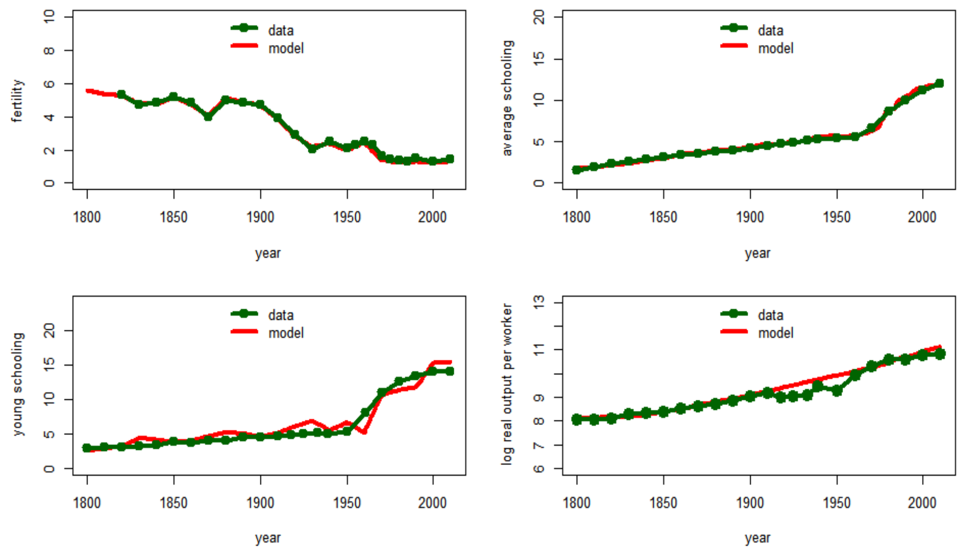

Figure 1.7: Fertility, schooling, real output per worker, young schooling (clockwise): Germany.

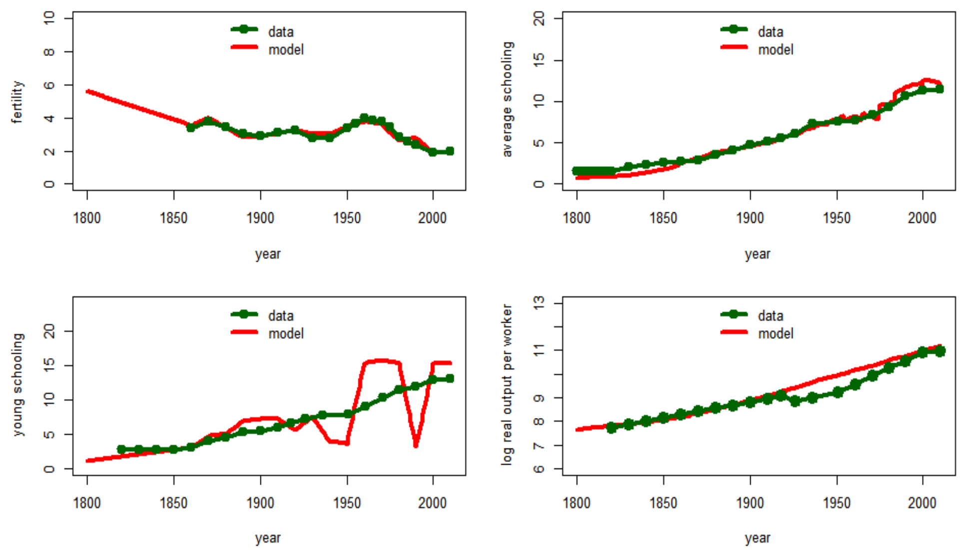

Figure 1.8: Fertility, schooling, real output per worker, young schooling (clockwise): Ireland.

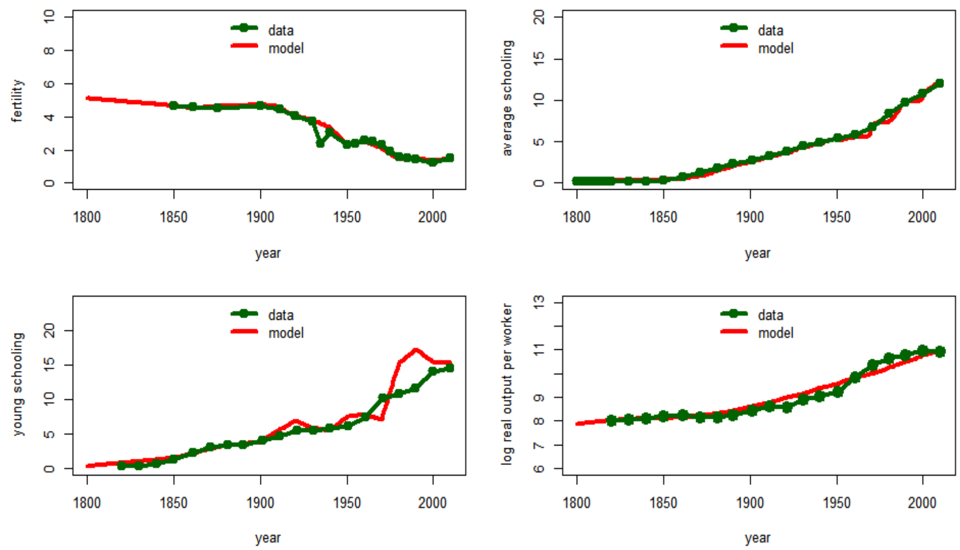

Figure 1.9: Fertility, schooling, real output per worker, young schooling (clockwise): Italy.

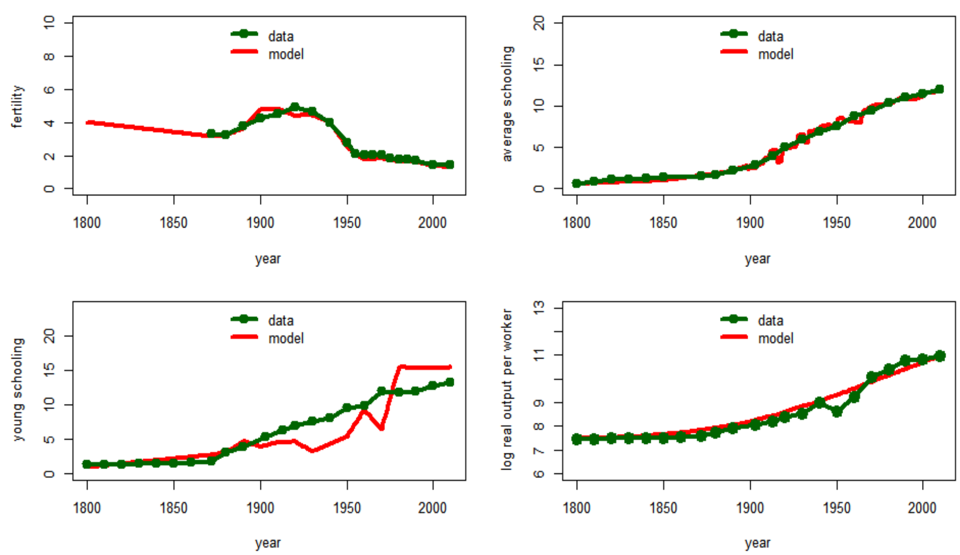

Figure 1.10: Fertility, schooling, real output per worker, young schooling (clockwise): Japan.

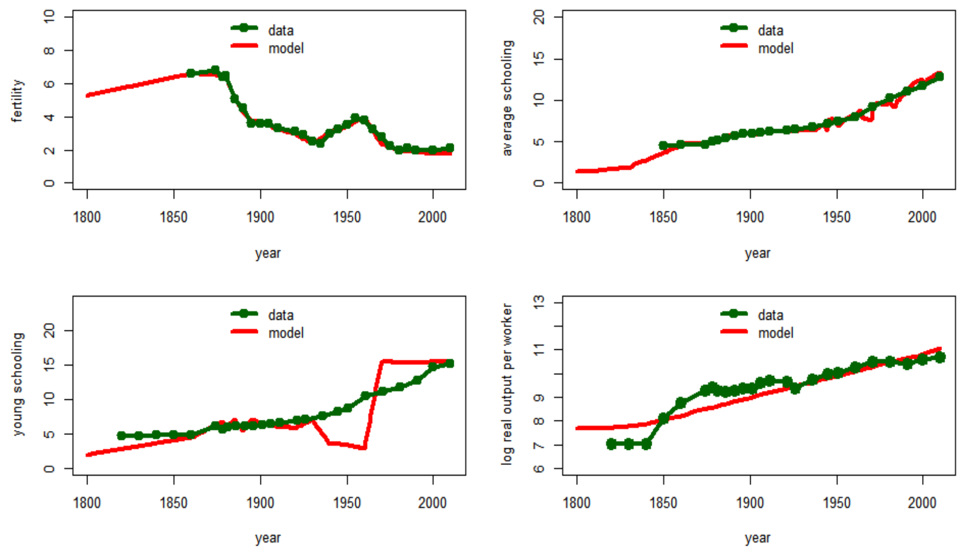

Figure 1.11: Fertility, schooling, real output per worker, young schooling (clockwise): Newzealand.

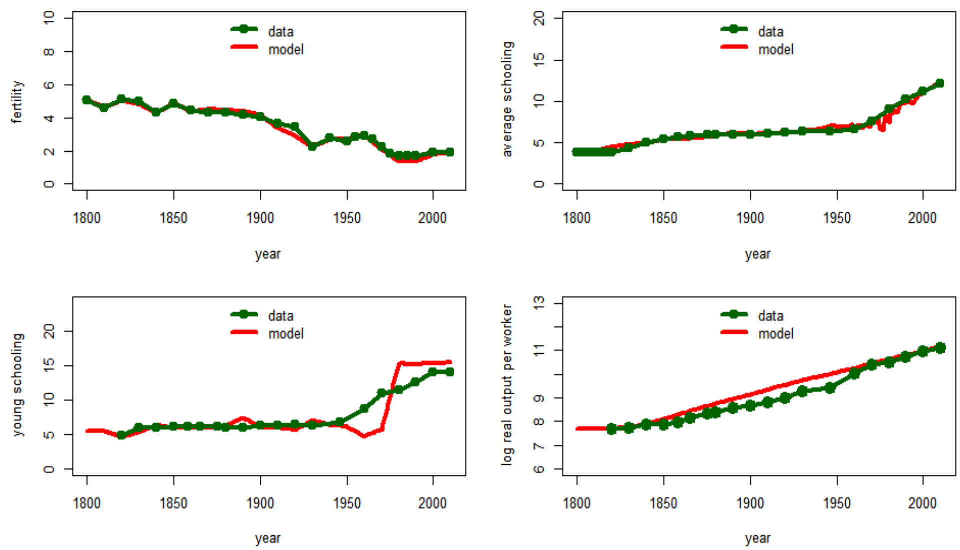

Figure 1.12: Fertility, schooling, real output per worker, young schooling (clockwise): Norway.

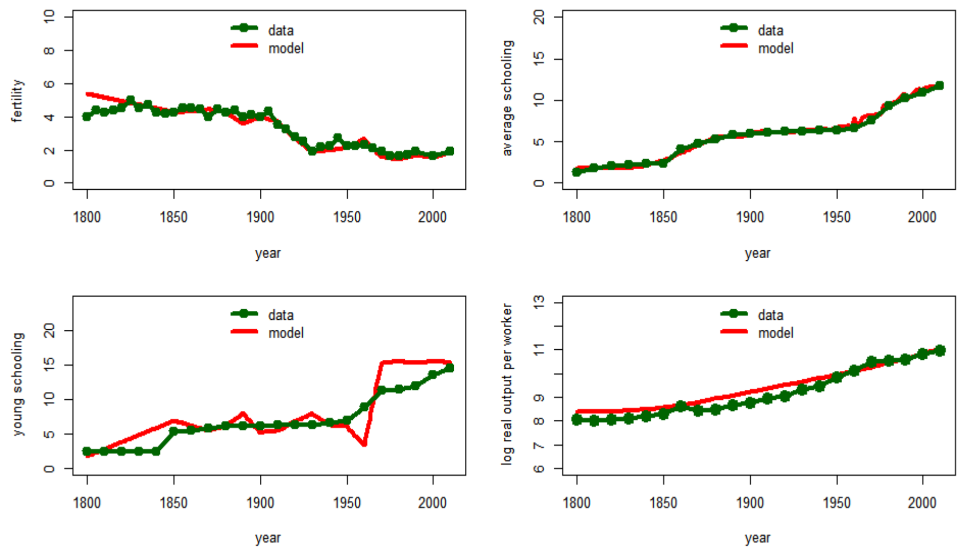

Figure 1.13: Fertility, schooling, real output per worker, young schooling (clockwise): Sweden.

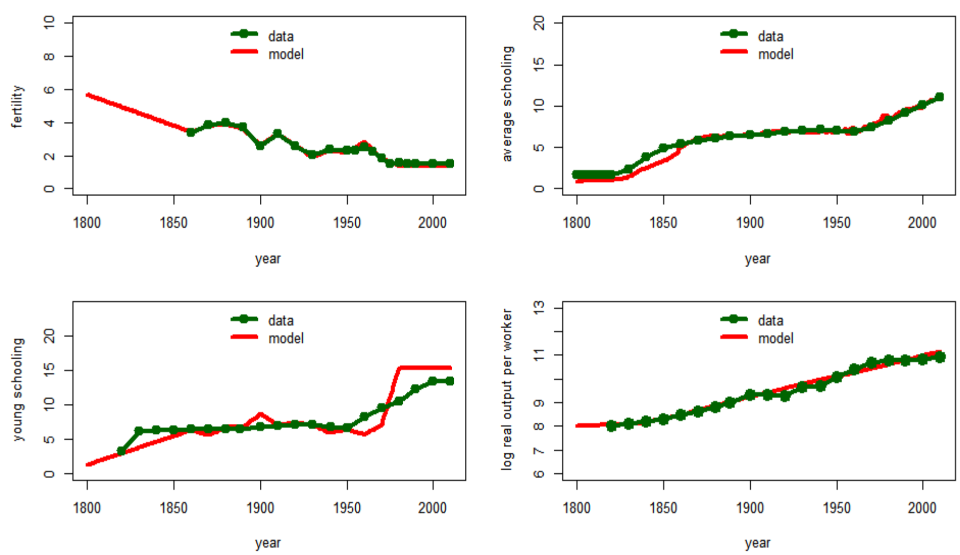

Figure 1.14: Fertility, schooling, real output per worker, young schooling (clockwise): Switzerland.

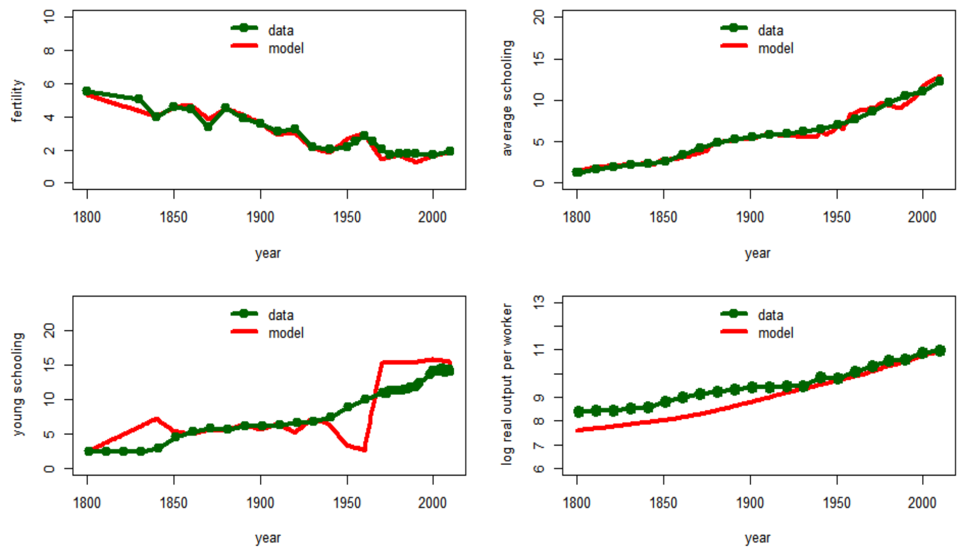

Figure 1.15: Fertility, schooling, real output per worker, young schooling (clockwise): The United Kingdom.

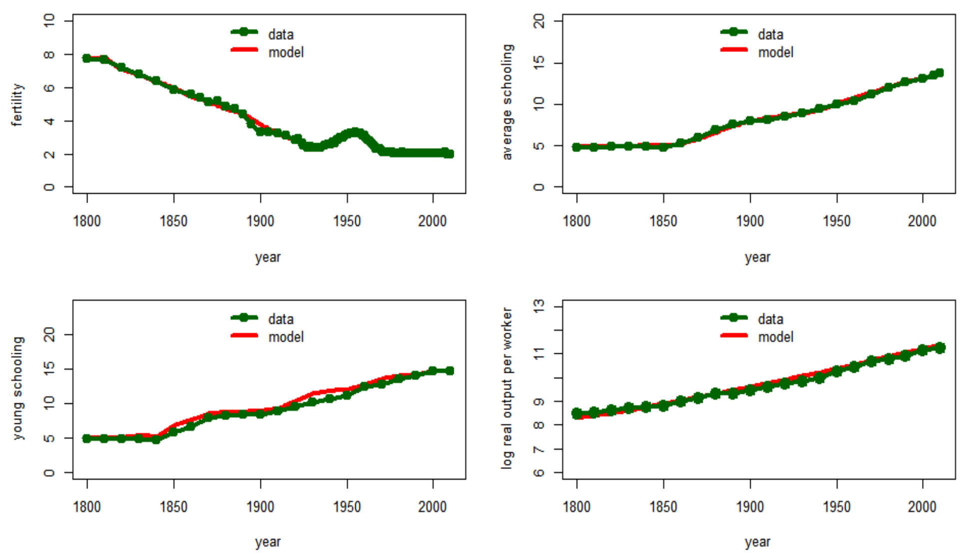

Figure 1.16: Fertility, schooling, real output per worker, young schooling (clockwise): United States.

Chapter 2

Quantifying Free Trade Agreements Using Machine Learning (with Dr. Scott Baier)

2.1 Introduction

The gravity model, often referred to as the workhorse model in international trade, has been widely used to study the effects of various determinants of trade flows across countries. Drawing from the analogy of physical science, Tinbergen (1962) first used gravity equation to evaluate the impact of free trade agreements (FTAs) on bilateral trade flows. Since Tinbergen (1962), numerous papers have studied the role of various determinants of trade flows, such as adjacency, common language, presence of a bilateral agreement, past colonial links, to name a few, cf., Head and Mayer (2014) . However, challenges abound in properly estimating the impact of free trade agreements on trade volumes. Broadly, there are two challenges that researchers ought to address to quantify the impact of FTAs accurately. We will now summarize the challenges and present the contribution of this paper in light of these challenges.

The first challenge is that trade policies can be potentially endogenous. Two countries are more likely to enter into a trade agreement if they are already significant trading partners. The possible reverse causality implies that trade policy variable is endogenous thereby making identification challenging. Another possible source of endogeneity is when trade policies are correlated to unmeasurable trade costs between the two countries, which may induce the two countries to "self-select" into a free trade agreement (see Baier and Bergstrand (2007) for a detailed analysis of the sources of endogeneity).

Several studies identify the issue of endogeneity and show that the estimates that do not allow for simultaneous determination of trade policy and trade flows are highly underestimated, cf., Trefler (1993), Lee and Swagel (1997), Baier and Bergstrand (2007). Trefler (1993) shows that when trade policy is modeled endogenously, by allowing for the simultaneous determination of imports and non-tariff barriers (NTBs) in the US manufacturing, the restrictive impact of NTBs increases tenfold. Lee and Swagel (1997) also find that the exogenous treatment of trade flows and the presence of FTA leads to an underestimation of the role of FTAs.

Prior to Baier and Bergstrand (2007), most studies that recognized endogeneity either used instrumental variable or control function techniques in cross-sectional data. Baier and Bergstrand (2007) show that the estimates obtained using cross-section instrumental variable and control function approaches are unstable in the presence of endogeneity. They show that using panel data with country-pair fixed effects accounts for endogeneity and leads to an unbiased estimation of the impact of FTAs. They find that an FTA will on average increase two member countries trade by about 86 percent after 15 years, six times the effect using OLS.

Another important consideration in properly estimating the impact of FTAs is that there exists an extensive heterogeneity across FTAs in terms of treaty design, coverage areas of trade policies, legal enforceability, and even the overall objectives. Consider, for example, the India-Bhutan Free Trade Agreement and the NAFTA. The former does not go beyond commitments in tariff liberalization in goods, whereas the latter covers commitments in a wide array of topics including goods and services liberalization, investment liberalization, environmental and labor standards, to name a few. The motive of signing free trade agreements can also differ across country pairs. Rosen (2004) provides evidence that the US-Israel Free Trade Agreement and the US-Jordan Free Trade Agreement are means of using trade policy to pursue foreign policy objectives.

Unlike endogeneity, capturing heterogeneity in trade agreements remains a challenge in the literature. The most common approach is to treat the existence of a free trade agreement between trading partners as an indicator variable to estimate the common average effect across agreements, cf., Baier and Bergstrand (2007), Anderson, Milot, and Yotov (2011), Anderson and Yotov (2016). This methodology, however, cannot take into account the fact that FTAs differ extensively in terms of the scope and the level of integration commitments between the parties.

The second approach involves what Kohl, Brakman, and Garretsen (2016) call a “specialist” approach, in which researchers examine the effect of individual FTAs on the members’ trade volumes. At best, researchers restrict themselves to a small number of trade agreements in a geographical region with a shorter time horizon. As such, the generalizations from these studies can be difficult, and the policy implications might be limited.

Baier, Bergstrand, and Feng (2014) provide the first evidence of the differential effects of trade agreements, which is not restricted to a geographical region. They categorize a large number of trade agreements based on their level of economic integration namely, non-reciprocal preferential trade agreements, reciprocal preferential trade agreements, free trade agreements, customs unions, common markets, and economic unions based on the traditional definition by Frankel, Stein, and Wei (1997). Using gravity model of trade flows, they show that more comprehensive trade agreements result in more trade creation.

The commitments in modern trade agreements, however, go far beyond tariff barriers or factor market integration and incorporate numerous policy areas that may affect the overall bilateral trade costs among member countries. For example, Horn, Mavroidis, and Sapir (2010) examine the content of 14 EC and 14 US trade agreements by going through the 28 agreements in their entirety and identify up to 52 policy areas included in the trade agreements signed by the EU and the US. This suggests that examining the differential effects of trade agreements requires a detailed analysis of these extensive set of policy areas.

Kohl, Brakman, and Garretsen 2016 is the only study that examines the coverage and restrictiveness of policy provisions in trade agreements. The authors of Kohl, Brakman, and Garretsen (2016), KBG hereafter, read through 296 trade agreements and quantify them based on the 17 policy areas they cover and the legal enforceability of these provisions. For each provision, a score of 0 is assigned if the document does not include the provision, 1 if it does include the provision, and 2 if it includes the provision and the provision is also legally enforceable. They then add up these scores and obtain a composite score for each trade agreement.

While highly informative and distinct in its approach, KBG study suffers from a few problems, and our paper by using a machine learning approach mitigates those problems. First, the assignment of 0, 1, and 2 seems quite ad-hoc, and a score of 2 does not necessarily imply that the provision is comprehensive. In KBG study, a score of 2 is assigned as long as a policy provision is covered and if there is a binding word such as "shall" or "must" present within the provision regardless of the specificity and the comprehensiveness of the provision. In this paper, we use techniques grounded in the distribution of words and phrases in trade agreements to examine the comprehensiveness of coverage areas and their legal enforceability. Second, with KBG approach, there must be many provisions which are in the border of being between 0, 1, and 2. Using machine learning techniques, we will be able to keep judgment calls at bay. Besides, we will be able to capture nuances that Kohl's approach might have missed. Third, KBG adds up scores across all provisions and obtains a composite score as a measure of "depth" for each agreement. Adding up scores across all provisions may not be prudent because not all provisions are trade-promoting; some provisions may be trade-restricting as well. Instead, the clustering techniques we use in this paper mitigates this problem. Finally, KBG study only examines the coverage and the enforceability of 17 trade policy areas. Our study examines 37 policy areas, more than twice the number of policy areas covered by KBG study.

In this paper, we proceed in three steps. First, we classify trade agreements into distinct clusters using k-means clustering, an unsupervised learning method. Second, we use multi-label classification, a supervised learning method, to examine the nature of each cluster in terms of the coverage and comprehensiveness of specific trade-policies of interest. It turns out that the groupings of trade agreements obtained from clustering in the first stage carry economic interpretation. The clustering exercise can separate shallow agreements from deep agreements quite well. Third, we then run the gravity model of trade regression and find evidence that the trade agreements that cover a wide range of trade-policy areas with high legal enforceability lead to the most substantial impact on trade flows. Our approach also allows us to identify the provisions that have the strongest impact on trade creation. These provisions are anti-dumping, capital mobility, competition laws, customs harmonization, dispute settlement framework, e-commerce, environment, export restrictions, import restrictions, institutional arrangements, intellectual property rights, freedom of transit, investment, labor, public procurement, sanitary and phytosanitary measures, liberalization of services trade, subsidies and countervailing measures technical barriers to trade, telecommunications liberalization as well as transparency. These results have far-reaching implications in shaping modern trade policy institutions.

The structure of the paper is as follows. Section 2.2 explains the data used in the analysis. Section 2.3 and 2.4 present discussions on clustering and classification, respectively. Section 2.5 and 2.6 discuss the gravity model of trade flows as applied in the context of our study. Section 2.7 provides the main empirical results and findings, and section 2.8 concludes.

2.2 Data

We use a total of 290 FTA texts in our analysis. The trade agreement documents, as well as data on relevant bilateral country pairs, come from Baier and Bergstrand (2017)[1]. The complete list of 290 trade agreements is listed in Appendix A. We remove annexes, protocols, and schedules and focus on the main body of the text to ensure that our clustering results are not driven by the mere presence of schedules, protocols, and annexes.

2.3 Clustering

The application of machine learning methods requires conversion of trade agreement texts to numerical vectors. We remove punctuation, symbols, numbers, and white spaces and segment trade agreements texts into single words and two-word phrases. We then count the frequency of a single word and two-word phrases within a trade agreement and normalize the frequency by the size of the document. Essentially, each trade agreement is uniquely represented by a vector of normalized frequencies of single words and two-word phrases. Appendix B discusses the methods in detail.

The goal of a cluster analysis is to find natural groupings of objects based on a number of object features. The first stage of clustering involves determining an "appropriate" number of clusters, and the second stage involves grouping the trade agreements into the identified number of clusters. We use one of the most widely used clustering techniques called k-means clustering algorithm, and apply standard procedures in choosing the proper number of clusters by minimizing the sum of squared errors between the empirical mean of a cluster and the objects assigned to that cluster over all clusters.

We will now formally introduce k-means algorithm. Let $X = \{x_i\}$, $x_i \in \mathbb{R}^{\mathbb{D}}$, be the set of

[1]The database is available at: https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/29762

trade agreements to be clustered into a set of K clusters, $C = \{c_k : k = 1, .., K\}$ where D is the number of features per trade agreement. Assume a-priori that there exists K clusters with cluster centers $\mu_1, \mu_2, ..., \mu_k \in \mathbb{R}^{\mathbb{D}}$. [2] K-means clustering solves

$$\arg\max_{C} \sum_{i=1}^{K} \sum_{x \in c_i} \|\mathbf{x} - \mu_{\mathbf{i}}\|^2. \tag{2.1}$$

The implementation of this optimization takes place in the following steps.

1 Choose an initial number of clusters, k.

2 Initialize cluster centers $\mu_1, \mu_2, ..., \mu_k$ arbitrarily.

3 Given the fixed cluster centers, choose optimal group assignment for each data point (trade agreements) x_i based on the closest cluster center.

4 Update $\mu_1, \mu_2, ..., \mu_k$ on the basis of group assignments of x_i.

5 Repeat steps 3 and 4 until convergence i.e. the centroids of the cluster do not move.

We repeat the above steps for different values of K and choose an appropriate number of clusters, $k*$.

2.3.1 Optimal Number of Clusters

We use two techniques commonly used in the literature to determine the appropriate number of clusters. The first method is known as "Elbow Method", where within groups sums of squares is plotted against the number of clusters. If the plot resembles an arm, then the "elbow" on the arm is the appropriate number of clusters. Figure **??** plots the within groups sum of squares against the number of clusters. The appropriate number of cluster as suggested by elbow method is anywhere between 5 to 7. It is not always the case that the plot resembles a perfect elbow as is the case with our analysis. Therefore, it is worth employing a second technique as a robustness exercise.

Gap statistic developed by Tibshirani, Walther, and Hastie (2001) is a more sophisticated approach to determine the optimal number of clusters. For all clusters, the gap statistic compares the total within-cluster variation for different values of k with their expected values under null reference

[2] Although we assume that there exists K centers, we will eventually update this based on the value of our loss function.

distribution of the data. The reference dataset is generated using Monte Carlo simulations where the reference data are obtained by uniformly sampling from a bounding rectangle of the original data. Therefore, the gap statistic for a given k is given by

$$Gap_{n(k)} = E_n^* log(W_k) - log(W_k), \tag{2.2}$$

where E_n^* denotes the expectation under a sample size n from the reference distribution, and W_k is the sum of within cluster variation across all k clusters. Figure **??** plots the gap-statistic against the number of cluster and the optimal number of clusters suggested by this approach is 6. Throughout the rest of the paper, we will present results for when trade agreements are split into both 5 and 6 clusters. The results are quite consistent across both cases.

2.3.2 Six Clusters vs Five Clusters

As we move from 6 clusters to 5 clusters, it is natural for a few trade agreements to change their cluster memberships. However, it would be a cause for concern if many agreements are switching their cluster memberships. Therefore, we run the correlation between the two sets of clusters. It turns out that the clusters are fairly stable. Table 2.1 shows the correlation matrix between the two sets of clusters. The own-cluster correlation is pretty high for all clusters, and the cross-cluster correlation is very low indicating that most of the trade agreements stay in their "natural" groups as we change the number of clusters. It is interesting to note that as we move from 6 clusters to 5 clusters, most trade agreements in the sixth cluster get merged into the fifth cluster. The cross-cluster correlation between the sixth cluster and the fifth cluster is quite high, and the cross-cluster correlation between the sixth cluster and the fifth cluster is very low.

2.3.3 Stability of Clustering Results

We ensure the stability of our clustering results by running 1000 iterations of k-means clustering in the cases of both 5 and 6 clusters. We then compute the correlation between the 1000 sets of clustering results against the base clustering result that we use in this paper. In the case of 5 clusters, the mean correlation is 0.9796, and the standard deviation is 0.0656. This indicates that clustering results are highly stable, and trade agreements do not change cluster membership erratically. The correlation is also quite high in the case of 6 clusters. The mean correlation is

0.9294 and the standard deviation of 0.1513. Therefore, this makes us confident that the k-means clustering is a promising way to find natural grouping in the trade agreements.

2.4 The Characteristics of the Identified Clusters

Thus far we have grouped trade agreements into their natural groupings. However, we have not yet presented any insights into the actual content of the trade agreements in each cluster. We now turn to a discussion of supervised learning method, which enables us to analyze the content of trade agreements in each cluster.

2.4.1 Supervised Learning: Classification

Supervised learning involves inferring the function from labeled training data. The training data comprises of a set of training examples and a label associated with each training example. Based on the training examples and the user-specified labels for these examples, supervised learning method infers the label for a test data. More formally, given training examples of the form $(x_1, y_1), ...(x_N, y_N)$ where x_i is a vector of features and y_i is an assigned label, the goal of a learning algorithm is to infer a function $g : X \to Y$ and thus to predict the output label for an unseen test sample.

In the context of our study, we train our model with instances of 37 trade policy provisions and allow the machine learning algorithm to "crawl" through each paragraph of the trade agreement to estimate the likelihood of the paragraph being about one or more than one policy area. The 37 policy areas we identify are Agriculture, Anti-Corruption, Anti-Dumping, Capital Mobility, Competition, Consumer Protection, Cooperation in Science and Technology, Customs Administration, Dispute Settlement, E-commerce, Education and Training Cooperation, Energy, Environment, Financial Cooperation, Freedom of Transit, Export Restrictions, Import Restrictions, Industrial Cooperation, Investment, Institutional Arrangements, Intellectual Property Rights, Investor-State Dispute Settlement, Labor, Money Laundering and Illicit Drugs, Public Procurement, Safeguard Procedures, Sanitary and Phyto-Sanitary Measures, Services, Small and Medium-Sized Enterprises, State Aid, State Trading Enterprises, Subsidies and Countervailing Measures, Technical Barriers to Trade, Telecommunications, and Transparency.

We train the model with examples of highly comprehensive and legally binding provisions for each policy provision. For policy provisions that are already covered in the WTO agreements, we seek for commitments above and beyond the WTO agreements. For policy provisions that are not already a part of the WTO agreements, we compare the provisions in all trade agreements and choose the most comprehensive legally binding provisions spanning all geographic regions. What we mean by legally binding is that the provision is very specific, the provision contains at least a restrictive word such as "shall" or "must", and the provision specifies the course of action in case either party deviates from the commitment listed in the provision. This course of action may be different than a generic dispute settlement present in the trade agreements. The following provision below is training example for provision on *Investment*:

National Treatment

1. Each Country shall accord to investors of the other country and to their investments treatment no less favourable than that it accords in like circumstances to its own investors and to their investments with respect to the establishment, acquisition, expansion, management, operation, maintenance, use, possession, liquidation, sale, or other disposition of investments (hereinafter referred to in this Chapter as investment activities). Each Country shall accord to investors of the other Country and to their investments treatment no less favourable than that it accords in like circumstances to investors of a third State and to their investments, with respect to investment activities. Article 9.6: Performance Requirements.

Neither Party may impose or enforce any of the following requirements, enforce any commitment or undertaking, in connection with the establishment, acquisition, expansion, management, conduct, operation, or sale or other disposition of an investment of an investor of a Party or of a non- Party in its territory to:

(a) export a given level or percentage of goods or services;

(b) achieve a given level or percentage of domestic content;

(c) purchase, use or accord a preference to goods produced in its territory, or to purchase goods from persons in its territory;

(d) relate in any way the volume or value of imports to the volume or value of exports or to the amount of foreign exchange inflows associated with such investment;

(e) restrict sales of goods or services in its territory that such investment produces or provides by relating such sales in any way to the volume or value of its exports or foreign exchange earnings;

(f) transfer a particular technology, production process or other proprietary knowledge to a person

in its territory;

(g) supply exclusively from the territory of the Party the goods that it produces or the services that it provides to a specific regional market or to the world market.

The above paragraph is highly comprehensive about investment commitments between the parties, and the language of the text is also highly enforceable. We feed training examples such as the one above for each of the 37 provisions and perform a multi-label classification on each paragraph of trade agreements to estimate the likelihood of the paragraph being about one or more than one categories. Essentially, we find a similarity measure between an untrained paragraph and the trained examples. Since a paragraph can be about more than just one policy domain, we also allow for multi-label classification with a technique called cross-training (Boutell et al. (2004)). The idea is to use paragraphs with multiple labels more than once during training. This implies that each training example can be a positive instance for more than one category during training. We then perform multi-label classification as 37 individual binary classification problems using a one-vs-rest strategy. This method has been proven to be effective in classifying multi-label images in the classification literature. So for each unit of analysis (i.e. a paragraph), we estimate the probability of the unit being relevant in one or more than one of the 37 categories. As is common in the literature, we put a threshold probability of 0.5 for the document to be relevant about the category at all McLaughlin and Sherouse (2016).

We experimented with the two popular classification algorithms: K-Nearest Neighbors and Random Forest Classifier. We evaluated the performance of both models based on macro F1-scores averaged across all classes, the details of which will be presented in the next subsection. This process reveals that the K-Nearest Neighbors performs superior to Random Forest Classifier. We use K-Nearest Neighbors with 5 neighbors and uniform weights to classify our trade agreement paragraphs. With this, a brief discussion of K-Nearest Neighbors classifier is thus warranted.

2.4.2 K-Nearest Neighbors Classifier

K-Nearest Neighbor Classifier is one of the most popular non-parametric classification algorithms used in many machine learning applications. Given training data $D = (x_1, y_1), ..., (x_N, y_N)$ and a positive integer k where x_i is a vector of features and y_i represents self-assigned labels for those set of features, the class prediction for a new test point x_0 involves identifying the K observations in the training data that are closest to x_0. So the conditional probability for class j is given as

$$Pr(Y = j|X = x) = \frac{1}{K}\sum_{\Omega} I(y_i = j), \tag{2.3}$$

where Ω represents the set of K observations closest to the x_0 (Friedman, Hastie, and Tibshirani (2001)).

2.4.3 Model Evaluation

We evaluate our models and choose the appropriate parameters by performing classification on 5-folds cross-validated data. Cross-Validation is a model validation technique to assess the generalizability of the classification results. The idea is that the training examples are further separated into training and test set, and the classification model is estimated only on the training set. The model is then used to predict the class of the test set.

The metric used to evaluate the classification model the macro F1-scores averaged across all classes. F1-scores is the harmonic average of precision and recall. In binary classification setting, precision is the percentage of selected items that are correct. So, precision is given by

$$\frac{True\ Positives}{True\ Positives + False\ Positives}. \tag{2.4}$$

Similarly, recall is defined as the percentage of correct items that are selected. So, recall is given by

$$\frac{True\ Positives}{True\ Positives + False\ Negatives}. \tag{2.5}$$

Then F1 score is given by

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision\ +\ Recall}. \tag{2.6}$$

This measure strikes a right balance between precision and recall. For our classification, we average F1 scores across all classes. Table 2.2 presents average F1 scores across all classes from 5 fold cross-validation for K-Nearest Neighbors and Random Forest Classifier with various parameter values. It can be seen that K-Nearest Neighbors Classifier performs better than Random Forest Classifier for every parameter value tested. Within K-Nearest Neighbors Classifier, the model with 5 neighbors and uniform weight performs the best classification, which is what we use for classification on our test data.

2.4.4 Provision Scores

Table 2.3 and Table 2.4 present the average provision score by cluster both in the cases of 6 and 5 clusters, respectively. In both tables, we re-arrange the clusters so that the first column represents the average provision scores for the shallowest cluster (least comprehensive) and the last column represents the average provision scores for the deepest cluster (most comprehensive). We will discuss more about these results against the backdrop of gravity regression results, which is the topic for next section.

2.5 Gravity Analysis with Cluster

In the second section, we clustered trade agreements in their natural groupings. In the subsequent section, we delved into the actual content of trade agreements belonging to each cluster. In this section, we combine the cluster membership information obtained from the first stage with gravity analysis of trade flows to obtain more insights into the heterogeneous impacts of FTAs on trade flows. The gravity regression results along with the provision-wise scores obtained from the supervised classification method will enable us to answer policy relevant questions such as what provisions or the set of provisions matters the most for trade flows and under what conditions.

2.6 Gravity Model

The gravity model, often referred to as the workhorse model in international trade, is widely used to study the effects of various determinants of country pairs' goods and factors flows. The gravity model uses the metaphor of Newton's Law of Gravitation and predicts that the trade flows between two countries is directly proportional to the product of their economic sizes and inversely proportional to the distance between their centers.[3] Tinbergen (1962) use gravity equation to eval-

[3] In Newton's Law of Universal Gravitation, the gravitational force between the two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

uate the effects of FTA dummy variables on trade. Despite its empirical success, the initial applications are a-theoretical. Anderson (1979) is the first study that lay the microeconomic foundations for the gravity equation under the assumption of CES identical CES preferences across countries and product differentiation by place of origin. The other studies that lay the theoretical groundwork are Bergstrand (1985), Eaton and Kortum (2002) and Anderson and Van Wincoop (2003).

Following Hummels (2001) and Baier and Bergstrand (2007), we follow the following gravity specification:

$$lnX_{ij,t} = \delta_{i,t} + \delta_{j,t} + \chi_{ij} + \sum_{k=1}^{K} FTA_{ij,t} * Cluster_k + \epsilon_{ij,t}, \tag{2.7}$$

where the $lnX_{ij,t}$ represents the logarithm of trade flows from country i to country j, $\delta_{i,t}$ is the vector of exporter-time fixed effects that captures any exporter-specific factors, $\delta_{j,t}$ is the vector of importer-time fixed effects that captures any importer-specific factors, and the vector χ_{ij} denotes time-invariant country pair fixed effects. The first two fixed effects also control for "multilateral resistance" first introduced by Anderson and van Wincoop (2003), where they argue that trade flows between two countries not only depend on the bilateral trade barriers but also on the trade resistance across all trading partners. As shown in Baier and Bergstrand (2007), the country pair fixed effects will take into account any possible endogeneity of free trade agreements. $FTA_{ij,t}$ is a dummy variable that takes a value of 1 if country i and country j have an FTA between them and a value of 0 otherwise. $Cluster_k$ is also a dummy variable and takes a value of 1 if $FTA_{ij,t}$ belongs to cluster k. K is the total number of clusters.

We run model specification (2.7) for two different sets of clustering results: one for 6 clusters as suggested by the gap statistic, and one for 5 clusters for robustness purposes.

2.7 Results and Discussion

After performing the above text mining exercises along with the gravity analysis of trade flows, we have a compelling way to answer the following questions. Do more comprehensive agreements result in more trade creation? What trade policy provision or what set of trade policy provisions matter the most for trade flows?

The first column of 2.5 shows the gravity trade regression results with 6 clusters, and Table 2.3 shows the average cluster scores for each provision in the case of 6 clusters. The gravity results indicate that the most comprehensive set of agreements, denoted by *deepest*, have the largest impact on trade flows. The coefficient for this cluster is about 0.83. This indicates that the most comprehensive trade agreements can increase two members' trade flows by up to 130% ($e^{0.83} = 2.30$).

The prominent provisions in the *deepest* cluster are anti-dumping, capital mobility, competition laws, customs harmonization, dispute settlement mechanism, e-commerce, environment, export restrictions, import restrictions, institutional arrangements, intellectual property rights, freedom of transit, investment, labor, public procurement, sanitary and phyto-sanitary measures, liberalization in services trade, subsidies and countervailing measures technical barriers to trade, telecommunications liberalization, and transparency.

The most comprehensive set of agreements remains the most influencing in terms of its impact on trade flows as we split trade agreements into only 5 clusters. The second column of Table 2.5 shows the gravity trade regression results with 5 clusters. The FTA coefficient is 0.58. The coefficient falls because as we decrease the number of clusters to 5, the trade agreements that are relatively less comprehensive, yet very similar, compared to the *deepest* cluster also get grouped into the same cluster. Nevertheless, the prominent provisions present in these clusters are the same as evidenced by high scores on the same set of provisions in Table 2.4.

The FTA coefficients for the next two most comprehensive set of trade agreements, denoted by *deeper* and *deep* in the first column of Table 2.5, are also high. The coefficients are 0.49 and 0.55, respectively. The coefficient is 0.57 in the case of 5 clusters. Examining the provision scores reveals that the same set of provisions that were present in the most comprehensive agreements are also dominantly present, albeit with relatively lower scores as expected.

Our results also indicate that shallow agreements, in general, have a meager impact in creating trade flows. The only exception is the shallowest group when we split the agreements into 6 clusters. The FTA coefficient is 0.41. But as we split the agreements into only 5 clusters, the coefficient for the shallowest group also falls. This indicates that grouping the trade agreements into 5 clusters may suffice.

2.7.1 Correlation of Provision Scores

We perform correlation on all of our provision scores simultaneously to understand how provisions co-occur in trade agreements. Figure 2.3 represents the correlation matrix heat map. Dark red cells represent a very high positive correlation whereas white cells represent zero correlation.

There are a few sets of provisions that stand out in terms of their high correlation. There is a strong evidence that the provisions on anti-corruption, e-commerce and environment tend to co-occur together. Similarly, the provisions on dispute settlement framework, investment, investor-state dispute settlement, public procurement, services and technical barriers to trade also tend to appear together. There are also some provisions for which the correlation with other any provision is almost zero. Provisions on measures on money laundering and illicit drugs, small and medium-sized enterprises, and cooperation on education and training do not have any discernible pattern to co-occur with other provisions.

2.8 Conclusion and Future Work

In this paper, we capture heterogeneity in free trade agreements using the tools of machine learning. The tools of machine learning allow us to quantify several features of trade agreements, including comprehensiveness and legal enforceability of provisions. First, we employ unsupervised learning techniques to categorize agreements into 5 to 6 clusters. Second, we use supervised learning techniques to analyze the content in each cluster in terms of the coverage of policy areas and the legal enforceability of those provisions. Finally, assuming that the trade flow effects are common across agreements within each cluster, we run the gravity model of trade flows to estimate the differential effects of free trade agreements. We find that more comprehensive agreements result in larger trade creation. In addition, we also identify the provisions that are generally the most successful in driving trade flows. The provisions are anti-dumping, capital mobility, competition, customs harmonization, dispute settlement mechanism, e-commerce, environmental standards, export restrictions, freedom of transit, import restrictions, institutional arrangements, intellectual property rights, investment, labor standards, public procurement, sanitary and phytosanitary measures, services, subsidies and countervailing measures, technical barriers to trade, telecommunications, and transparency.

One crucial question that remains unanswered is whether the benefits of incorporating the "trade-inducing" provisions that we identify in our paper extend to an arbitrarily chosen pair of countries. Under what set of country characteristics can a randomly chosen pair be expected to benefit the most from including the same set of provisions in a newly minted trade agreement? In future work, we will address this question by simultaneously exploring the relationship between country characteristics, policy provisions, and trade flows.

Another question that is worth exploring is how the content and legal enforceability of specific provision or a set of provisions affect trade flows in the relevant sector. For example, what aspects of provisions related to automobiles affect the trade flows in automobiles across the member countries? Is one method of rules of origin determination for tariff concessions more restrictive and trade hindering than the other? The answers to such questions may also be obtained by combining a standard empirical model of trade with the text analysis of provisions enabled by machine learning. I believe that this stream of research will have far-reaching implications in shaping modern trade institutions.

Table 2.1: Correlation between the two sets of clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	**0.92**	-0.23	-0.18	-0.32	-0.24
Cluster 2	-0.21	**0.86**	-0.13	-0.24	-0.17
Cluster 3	-0.16	-0.12	**0.96**	-0.24	-0.13
Cluster 4	-0.	-0.31	-0.23	**0.92**	-0.32
Cluster 5	-0.15	-0.12	-0.01	-0.24	**0.61**
Cluster 6	-0.18	0.05	-0.11	-0.28	**0.65**

This table shows the correlation matrix between the two sets of clusters. The highlighted values represent the own-cluster correlation. A high own-cluster correlation value represents that the trade agreements do not change their groups drastically as we split trade agreements from 6 clusters to 5 clusters.

Table 2.2: Macro F1-Scores by Model: Classification

Model	Average Macro F1 scores
K-Nearest Neighbors, weight= uniform and K= 7	0.8366
	(0.0696)
K-Nearest Neighbors, weight= distance and K= 7	0.8160
	(0.0704)
K-Nearest Neighbors, weight= uniform and K = 6	0.8243
	(0.0632)
K-Nearest Neighbors, weight= distance and K = 6	0.8214
	(0.0635)
K-Nearest Neighbors, weight= uniform and K= 5	0.8373
	(0.0649)
K-Nearest Neighbors, weight= distance and K = 5	0.8123
	(0.0649)
Random Forest Classifier, number of trees=5	0.5712
	(0.0704)
Random Forest Classifier, number of trees=10	0.5388
	(0.0635)
Random Forest Classifier, number of trees=15	0.5991
	(0.0628)

1. F1 scores are averaged across all classes based on 5-fold Cross Validation.

2. For K-Nearest Neighbor model, represents the number of neighbors used.The weight parameter of distance indicates that within a class, closer neighbors will have a greater influence than neighbors which are farther away.

Table 2.3: Provision Scores for Number of Clusters = 6

Provisions	Shallowest	Shallow	Moderate	Deep	Deeper	Deepest
Agriculture	0.25	0.67	0.74	0.67	0.54	0.44
Anti-Corruption	0.00	0.00	0.00	0.39	0.05	0.27
Anti-Dumping	0.08	0.50	0.28	0.33	0.63	0.79
Capital Mobility	0.11	0.58	0.22	0.87	0.78	0.93
Competition	0.17	0.45	0.55	0.54	0.69	0.71
Consumer Protection	0.00	0.13	0.02	0.28	0.16	0.16
Customs Administration	0.30	0.63	0.10	0.83	0.83	0.98
Dispute Settlement	0.28	0.73	0.36	0.81	0.88	0.98
E-commerce	0.00	0.14	0.02	0.52	0.35	0.60
Education & Training Cooperation	0.02	0.21	0.03	0.17	0.32	0.24
Energy	0.05	0.16	0.03	0.15	0.16	0.19
Environment	0.05	0.21	0.08	0.65	0.39	0.63
Export Restrictions	0.69	0.91	0.88	0.93	0.88	0.95
Financial Cooperation	0.05	0.27	0.04	0.20	0.23	0.11
Freedom of Transit	0.20	0.31	0.03	0.57	0.36	0.59
Import Restrictions	0.70	0.89	0.86	0.92	0.90	0.97
Industrial Cooperation	0.13	0.29	0.05	0.23	0.33	0.24
Institutional Arrangements	0.37	0.72	0.94	0.66	0.78	0.65
Intellectual Property Rights	0.30	0.54	0.80	0.68	0.65	0.77
Investment	0.11	0.34	0.10	0.71	0.57	0.85
Investor-State Dispute Settlement	0.05	0.27	0.05	0.76	0.57	0.70
Labor	0.03	0.45	0.06	0.83	0.76	0.92
Money Laundering/Illegal Drugs	0.01	0.21	0.04	0.12	0.13	0.16
Political Dialogue	0.16	0.34	0.21	0.27	0.22	0.11
Public Procurement	0.06	0.40	0.48	0.75	0.55	0.81
Safeguard Procedures	0.20	0.71	0.97	0.68	0.66	0.61
Sanitary &Phyto-Sanitary Measures	0.13	0.39	0.41	0.65	0.58	0.77
Science and Technology Cooperation	0.12	0.23	0.05	0.36	0.49	0.40
Services	0.09	0.43	0.03	0.72	0.67	0.93
Small & Medium Sized Enterprises	0.03	0.20	0.05	0.13	0.23	0.16
State Aid	0.17	0.40	0.53	0.49	0.42	0.36
State Trading Enterprises	0.21	0.44	0.72	0.63	0.47	0.59
Subsidies and Countervailing Measures	0.08	0.44	0.40	0.45	0.47	0.47
Technical Barriers to Trade	0.21	0.59	0.36	0.68	0.71	0.88
Telecommunications	0.01	0.16	0.00	0.68	0.32	0.83
Transparency	0.00	0.51	0.06	0.90	0.86	0.93
Transportation Infrastructure	0.02	0.16	0.03	0.18	0.21	0.16

Notes: Table reports the average provision scores across trade agreements within each cluster in the case of 6 clusters. Low provision scores across all clusters implies that the provision is only prevalent in a few trade agreements.

Table 2.4: Provision Scores for Number of Clusters = 5

Provisions	Shallowest	Shallow	Moderate	Deep	Deepest
Agriculture	0.25	0.66	0.70	0.67	0.49
Anti-Corruption	0.00	0.02	0.00	0.37	0.15
Anti-Dumping	0.2	0.79	0.79	0.60	0.83
Capital Mobility	0.09	0.55	0.25	0.87	0.89
Competition	0.19	0.49	0.51	0.57	0.70
Consumer Protection	0.00	0.13	0.02	0.28	0.16
Customs Administration	0.26	0.69	0.14	0.85	0.90
Dispute Settlement	0.27	0.78	0.36	0.82	0.92
E-commerce	0.00	0.18	0.02	0.56	0.44
Education and Training Cooperation	0.02	0.21	0.04	0.16	0.30
Energy	0.03	0.13	0.06	0.17	0.17
Environment	0.05	0.22	0.09	0.65	0.51
Export Restrictions	0.69	0.91	0.85	0.92	0.93
Financial Cooperation	0.02	0.25	0.06	0.19	0.19
Freedom of Transit	0.17	0.29	0.06	0.59	0.48
Import Restrictions	0.69	0.91	0.85	0.92	0.93
Industrial Cooperation	0.09	0.24	0.10	0.24	0.30
Institutional Arrangements	0.34	0.75	0.90	0.64	0.73
Intellectual Property Rights	0.28	0.57	0.77	0.70	0.70
Investment	0.10	0.34	0.12	0.74	0.70
Investor-State Dispute Settlement	0.04	0.29	0.07	0.78	0.62
Labor	0.03	0.46	0.06	0.84	0.87
Political Dialogue	0.14	0.29	0.22	0.25	0.19
Public Procurement	0.02	0.64	0.58	0.78	0.78
Money Laundering/Illicit Drugs	0.00	0.19	0.05	0.11	0.15
Small and Medium Sized Enterprises	0.03	0.19	0.05	0.15	0.20
Sanitary and Phyto-Sanitary Measures	0.13	0.45	0.38	0.68	0.64
State Trading Enterprises	0.19	0.45	0.68	0.63	0.52
Safeguard Procedures	0.21	0.71	0.90	0.66	0.65
Science & Technology Cooperation	0.10	0.27	0.07	0.40	0.43
Services	0.07	0.44	0.10	0.74	0.80
State Aid	0.16	0.39	0.51	0.48	0.40
Subsidies & Countervailing Measures	0.08	0.48	0.38	0.44	0.45
Technical Barriers to Trade	0.23	0.64	0.34	0.71	0.78
Telecommunications	0.01	0.18	0.00	0.70	0.56
Transparency	0.00	0.60	0.06	0.91	0.88
Transport Infrastructure	0.01	0.15	0.04	0.18	0.19

Notes: Table reports the average provision scores across trade agreements within each cluster in the case of 6 clusters. Low provision scores across all clusters implies that the provision is only prevalent in a few trade agreements.

Table 2.5: Gravity Regressions with Different Sets of Clusters

Cluster Type	ln(Trade Flows)	ln(Trade Flows)
Shallowest	0.41***	0.28***
	(0.03)	(0.03)
Shallow	0.24***	0.28***
	(0.02)	(0.02)
Moderate	0.25***	0.31***
	(0.03)	(0.02)
Deep	0.55***	0.57***
	(0.03)	(0.03)
Deeper	0.49***	-
	(0.04)	
Deepest	0.83***	0.58***
	(0.06)	(0.03)
$\bar{R}^2$	0.8396	0.8395
N	612,182	612,182

Table reports the coefficients on the gravity regression with country pair fixed effects, country-import fixed effects, and country-export fixed effects. *** 1%, ** 5%, * 10%.

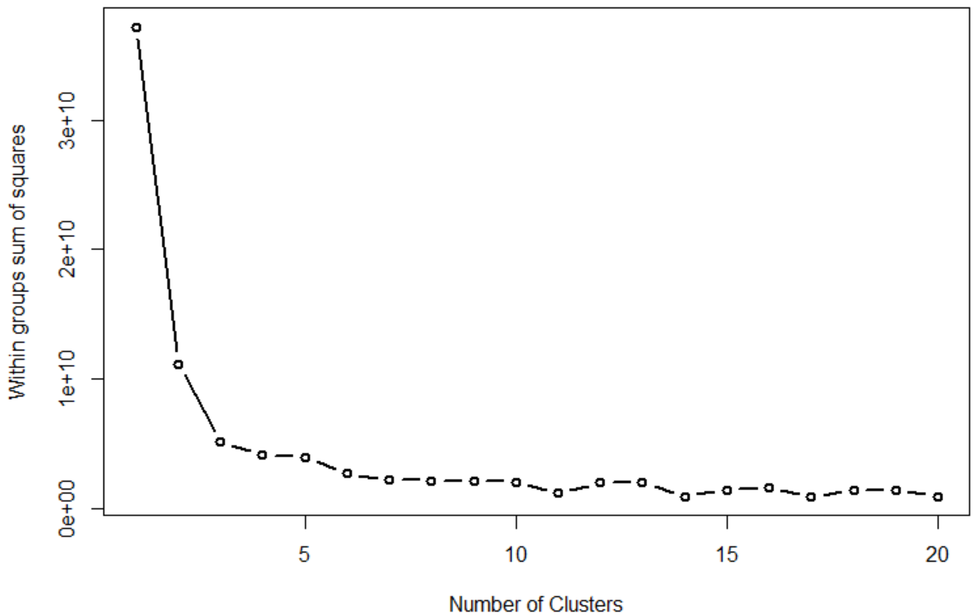

Figure 2.1: Within Groups Sums of Squares

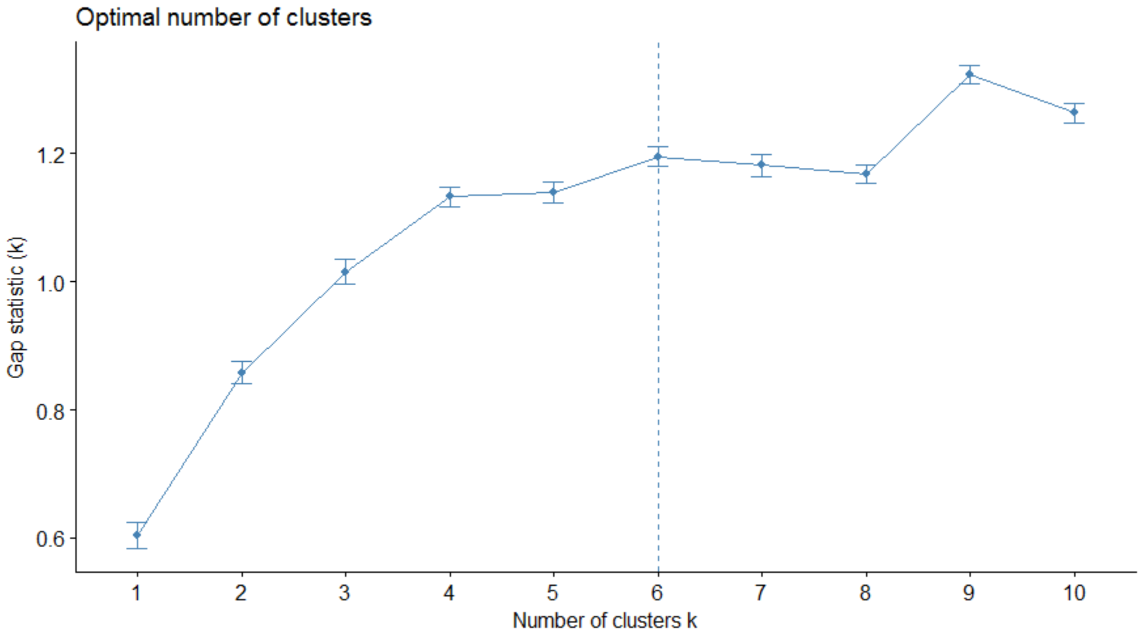

Figure 2.2: Gap-Statistic

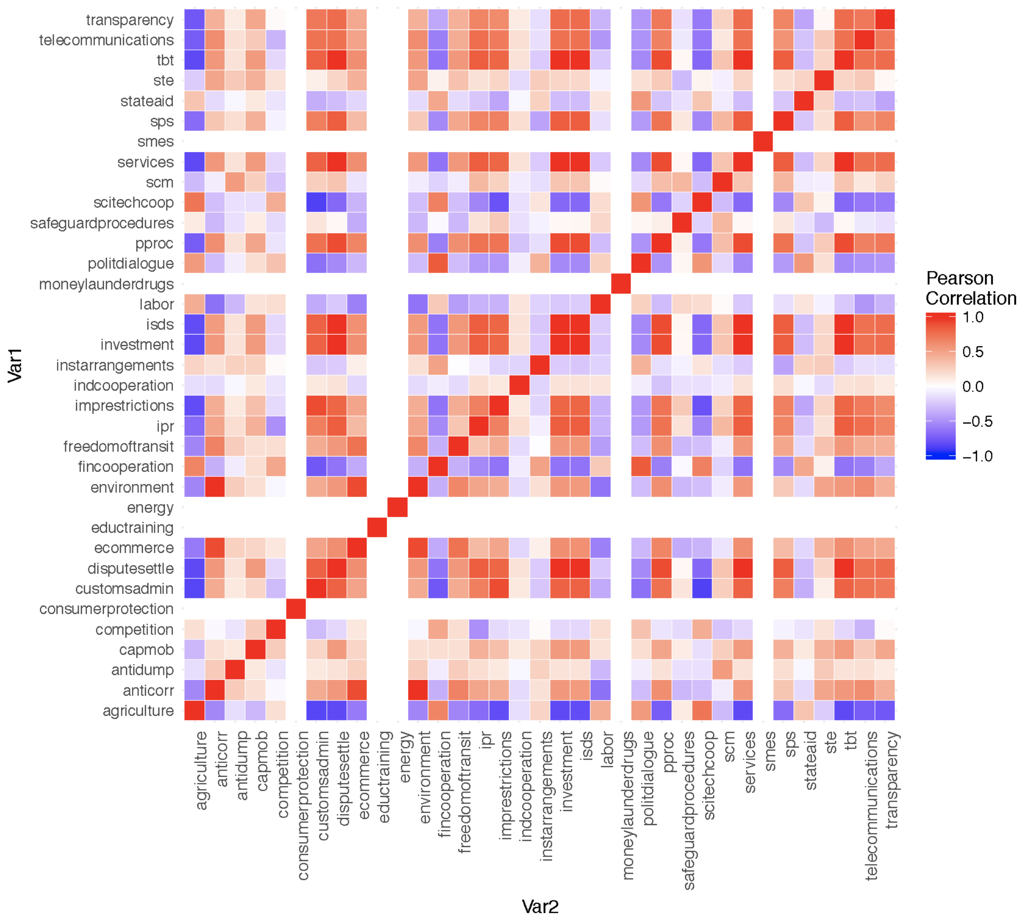

Figure 2.3: Heatmap

Appendices

Appendix A Time Series of Rent (r), κ, ν, and Mortality (δ)

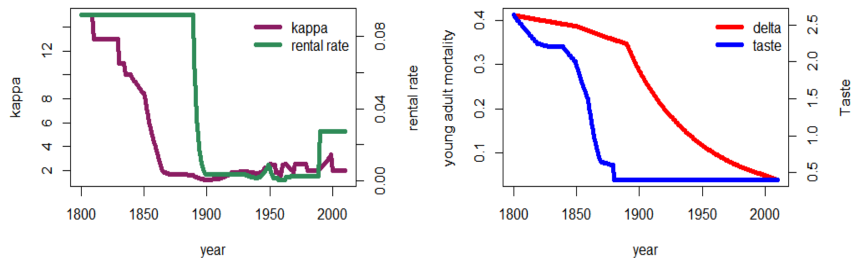

Figure 4: a) kappa and rental rate (left) b) delta and taste: Austria.

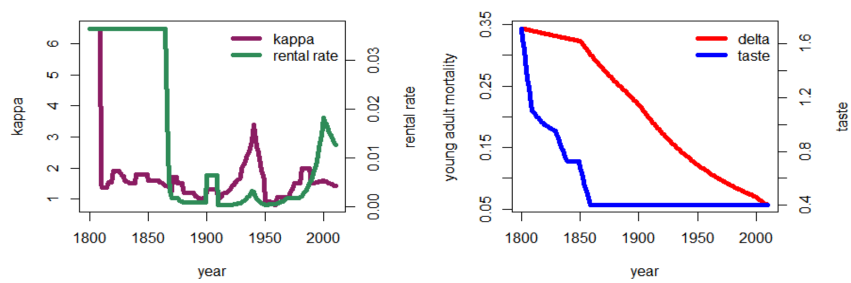

Figure 5: a) kappa and rental rate (left) b) delta and taste: Belgium.

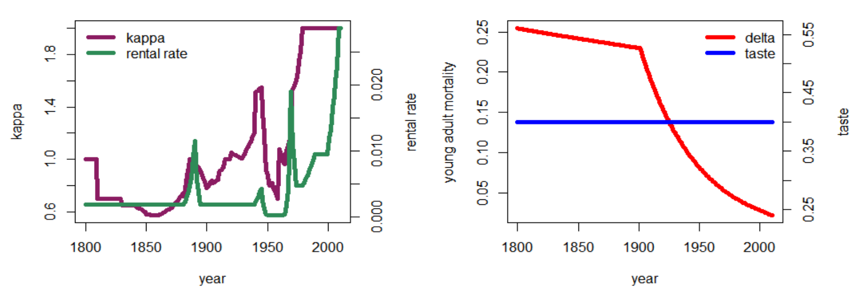

Figure 6: a) kappa and rental rate (left) b) delta and taste: Canada.

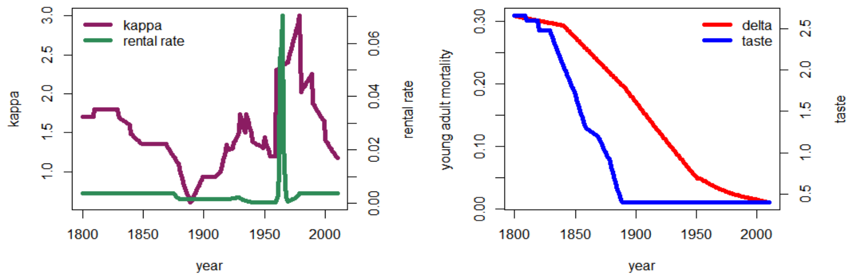

Figure 7: a) kappa and rental rate (left) b) delta and taste: Denmark.

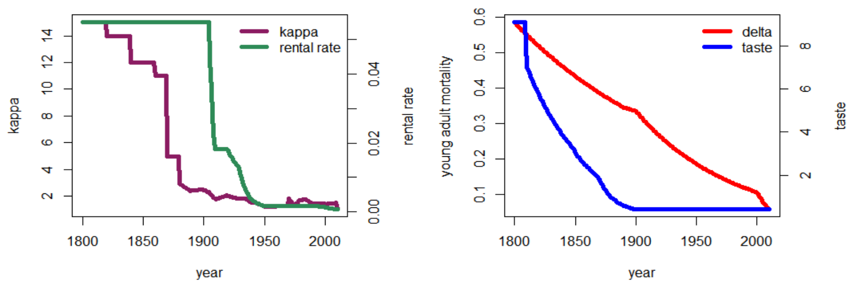

Figure 8: a) kappa and rental rate (left) b) delta and taste: Finland.

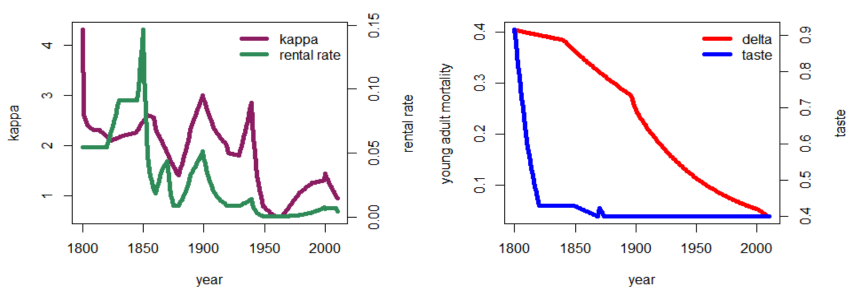

Figure 9: a) kappa and rental rate (left) b) delta and taste: France.

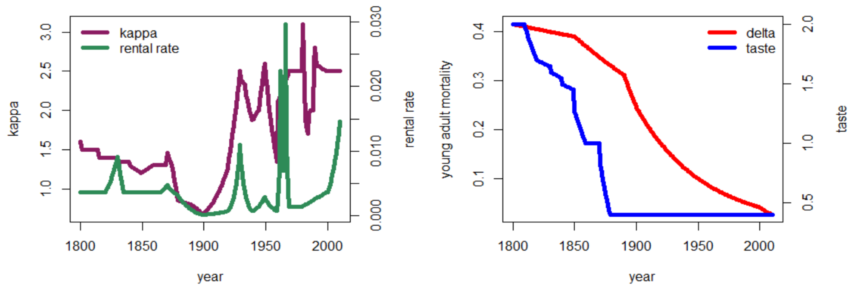

Figure 10: a) kappa and rental rate (left) b) delta and taste: Germany.

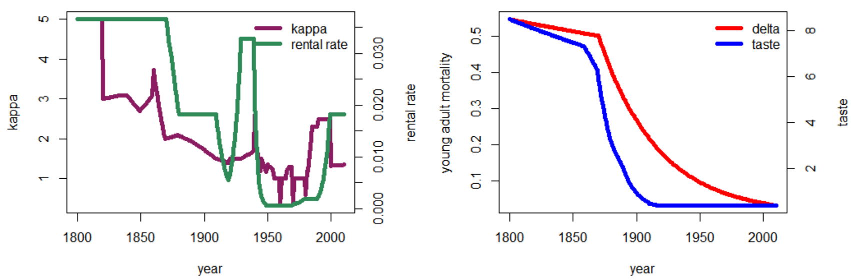

Figure 11: a) kappa and rental rate (left) b) delta and taste: Ireland.

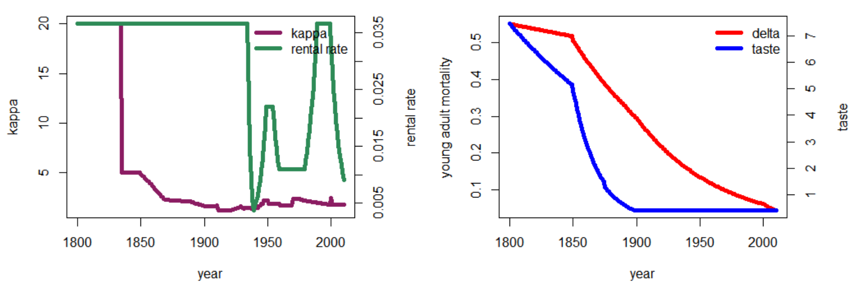

Figure 12: a) kappa and rental rate (left) b) delta and taste: Italy.

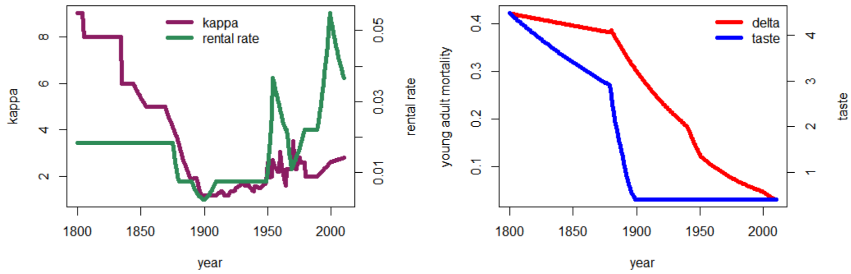

Figure 13: a) kappa and rental rate (left) b) delta and taste: Japan.

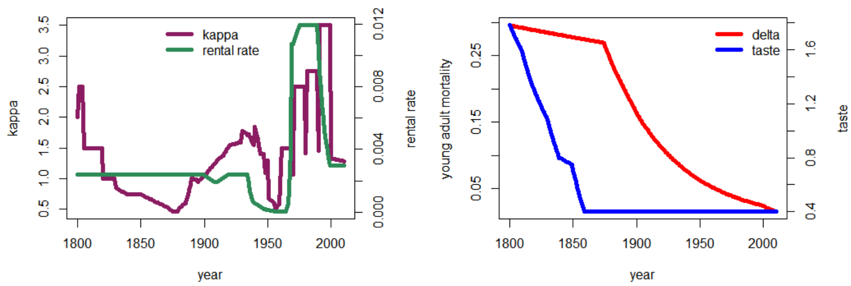

Figure 14: a) kappa and rental rate (left) b) delta and taste: New Zealand.

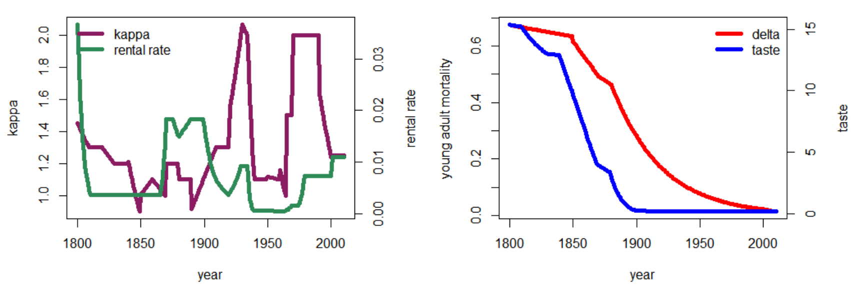

Figure 15: a) kappa and rental rate (left) b) delta and taste: Norway.

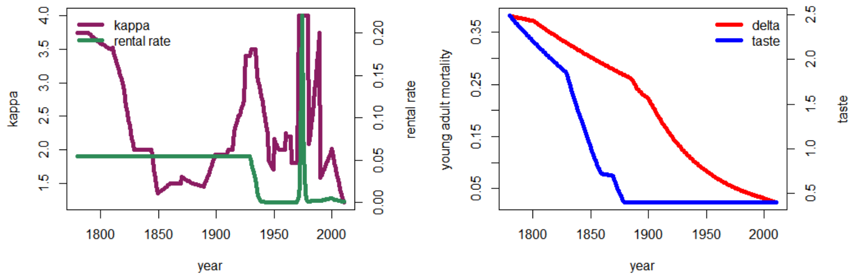

Figure 16: a) kappa and rental rate (left) b) delta and taste: Sweden.

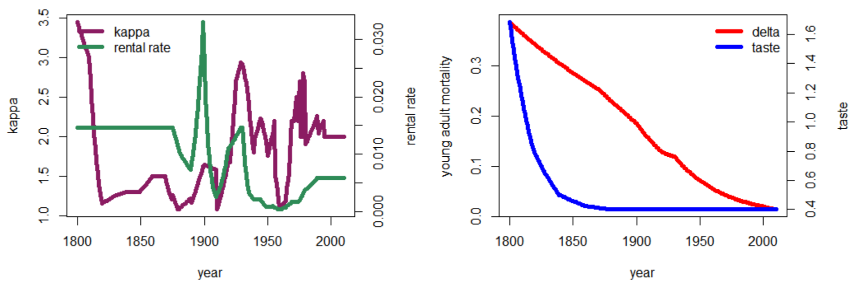

Figure 17: a) kappa and rental rate (left) b) delta and taste: Switzerland.

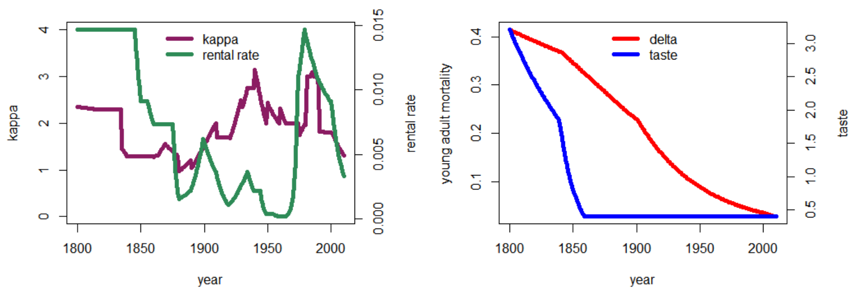

Figure 18: a) kappa and rental rate (left) b) delta and taste: United Kingdom.

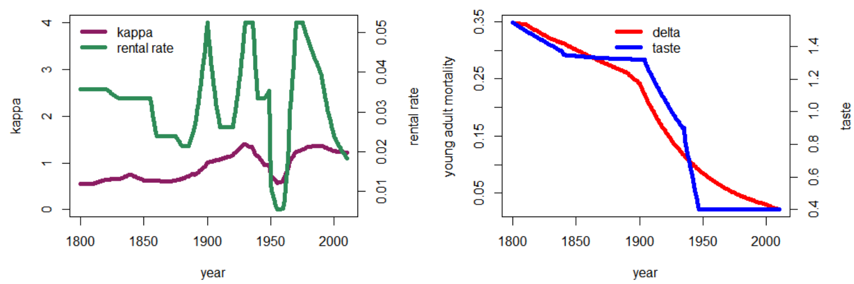

Figure 19: a) kappa and rental rate (left) b) delta and taste: United States.

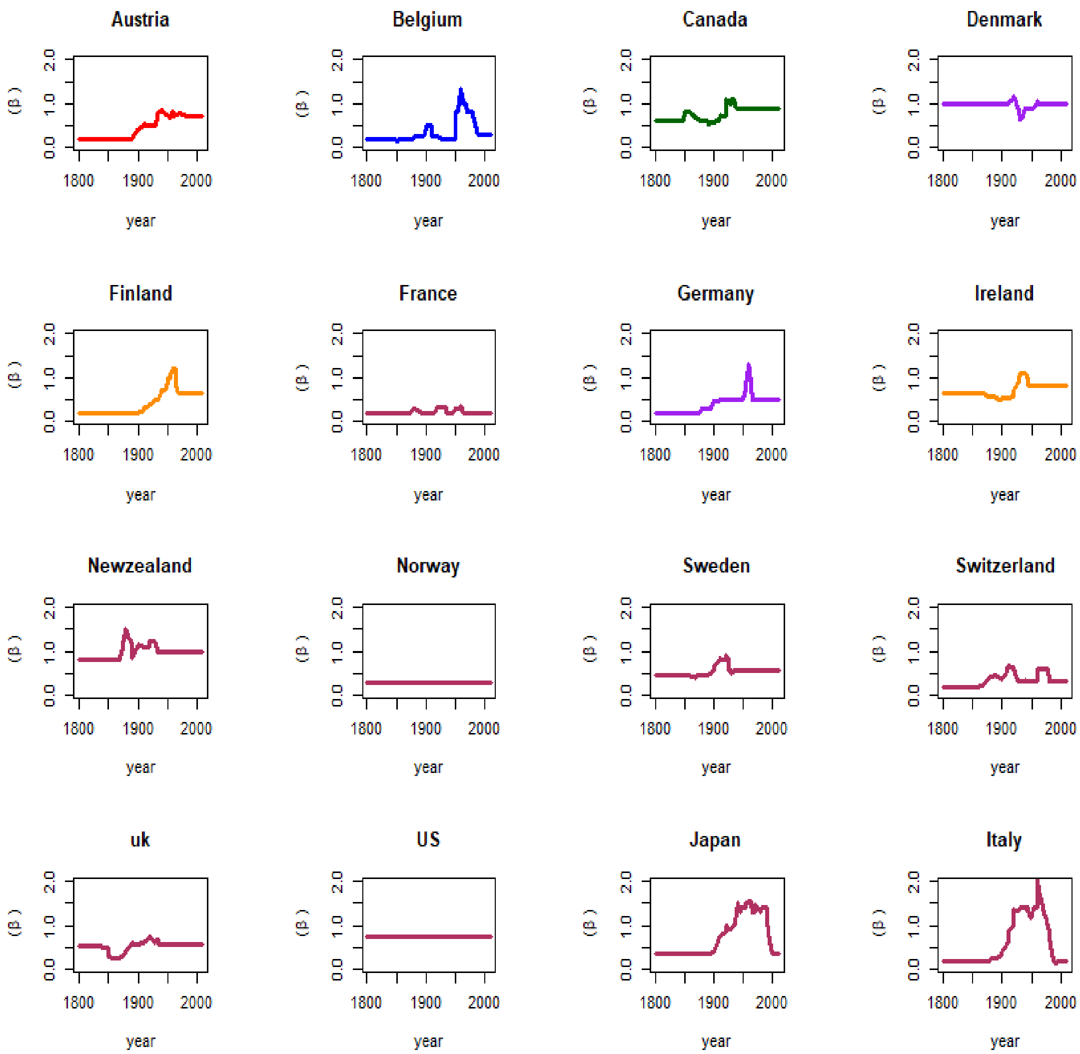

Figure 20: Parameter β_t across countries

Appendix B From Text Documents to Numerical Feature Vectors

For either clustering or classification analysis, the text documents will first need to be converted to a vector of real numbers. We follow a three step procedure that is commonly employed in natural language processing literature to transform text documents to numerical feature vectors. The first step involves assigning integer identification for each word or a two word combination, commonly referred to as tokenization. The trade agreement documents were tokenized using unigram (single word) and bigram counts (two word phrases). The words for tokenization are defined as sequences of two or more alphabetic characters, excluding stop words such as pronouns, articles, and prepositions that do not carry much meaning in differentiating one set of documents from other. We also remove punctuation, numbers and white spaces. The second step is to count the number of occurrences of these tokens for each document in the collection of document commonly referred to as corpus. The final step is to normalize each document to have a feature matrix of fixed size and to weight tokens that occur in the majority of documents with diminishing importance. We use tf-idf scheme developed by Salton and McGill (1983) to obtain weights for each token.

To calculate tf-idf values of a token in each document, we multiply the term frequency of that token by its idf component. This frequency is given by

$$idf(t) = log\frac{(1+n_d)}{1+df(d,t)} + 1, \tag{8}$$

where n_d is the total number of documents, and df(d,t) is the number of documents that contain term t. The tf-idf vectors thus obtained are then normalized by the Euclidean norm and is given by,

$$v_{norm} = \frac{v}{\sqrt{(v_1)^2 + (v_2)^2 + ... + (v_n)^2}}. \tag{9}$$

After completing this 3-step procedure, we obtain a features matrix X, whose row consists of tf-idf values of all possible tokens for each document in our corpus. Each row is also normalized to have a unit norm which is needed to account for the fact that documents in our corpus are of variable length. If we were to not normalize each row to a unit norm, a longer document will have higher term frequencies and thus higher tf-idf values than a shorter document.

Bibliography

Albanesi, Stefania, and Claudia Olivetti. 2014. "Maternal health and the baby boom". *Quantitative Economics* 5 (2): 225–269.

Anderson, James E, Catherine A Milot, and Yoto V Yotov. 2011. *The incidence of geography on canada's services trade*. Tech. rep. National Bureau of Economic Research.

Anderson, James E, and Eric Van Wincoop. 2003. "Gravity with gravitas: a solution to the border puzzle". *American economic review* 93 (1): 170–192.

Anderson, James E, and Yoto V Yotov. 2016. "Terms of trade and global efficiency effects of free trade agreements, 1990–2002". *Journal of International Economics* 99:279–298.

Baier, Scott L, and Jeffrey H Bergstrand. 2007. "Do free trade agreements actually increase members' international trade?" *Journal of international Economics* 71 (1): 72–95.

Baier, Scott L, Jeffrey H Bergstrand, and Matthew Clance. 2015. "Heterogeneous economic integration agreement effects".

Baier, Scott L, Jeffrey H Bergstrand, and Michael Feng. 2014. "Economic integration agreements and the margins of international trade". *Journal of International Economics* 93 (2): 339–350.

Becker, Gary S. 1960. "An economic analysis of fertility". In *Demographic and economic change in developed countries*, 209–240. Columbia University Press.

Becker, Gary S, and Robert J Barro. 1988. "A reformulation of the economic theory of fertility". *The quarterly journal of economics* 103 (1): 1–25.

Becker, Gary S, and H Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children". *Journal of political Economy* 81 (2, Part 2): S279–S288.

Becker, Gary S, Kevin M Murphy, and Robert Tamura. 1990. "Human capital, fertility, and economic growth". *Journal of political economy* 98 (5, Part 2): S12–S37.

Bergstrand, Jeffrey H. 1985. "The gravity equation in international trade: some microeconomic foundations and empirical evidence". *The review of economics and statistics*: 474–481.

Boutell, Matthew R, Jiebo Luo, Xipeng Shen, and Christopher M Brown. 2004. "Learning multi-label scene classification". *Pattern recognition* 37 (9): 1757–1771.

Caldwell, John C, and K Srinivasan. 1984. "New data on nuptiality and fertility in China". *Population and Development Review*: 71–79.

Castles, Lance. 1975. "Sources for the population history of northern Sumatra". *Masyarakat Indonesia Tahun Ke* 2 (2): 189–208.

Chesnais, Jean-Claude, et al. 1992. "The demographic transition: Stages, patterns, and economic implications". *OUP Catalogue*.

Coale, Ansley J. 1989. "Demographic transition". In *Social economics*, 16–23. Springer.

Collver, Andrew. 1965. "Birth rates in Latin America: New estimates of historical trends and fluctuations". In *Birth rates in Latin America: new estimates of historical trends and fluctuations.*

Concepcion, Mercedes B. 1977. "The population of the Philippines". *National Population Monographs. CICRED Series.*

Doepke, Matthias, Moshe Hazan, and Yishay D Maoz. 2015. "The baby boom and World War II: A macroeconomic analysis". *The Review of Economic Studies* 82 (3): 1031–1073.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, geography, and trade". *Econometrica* 70 (5): 1741–1779.

Economic Intergration Agreements Database. https://www3.nd.edu/ jbergstr/.

Frankel, Jeffrey A, Ernesto Stein, and Shang-Jin Wei. 1997. *Regional trading blocs in the world economic system.* Peterson Institute.

Frankel, Jeffrey, Ernesto Stein, and Shang-Jin Wei. 1995. "Trading blocs and the Americas: The natural, the unnatural, and the super-natural". *Journal of development economics* 47 (1): 61–95.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The elements of statistical learning.* Vol. 1. 10. Springer series in statistics New York, NY, USA:

Galor, Oded, and Omer Moav. 2002. "Natural selection and the origin of economic growth". *The Quarterly Journal of Economics* 117 (4): 1133–1191.

Galor, Oded, and David N Weil. 2000. "Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond". *American economic review* 90 (4): 806–828.

— . 1998. *Population, Technology, and Growth: From the Malthusian Regime to the Demographic Transition and Beyond.* Tech. rep. National Bureau of Economic Research.

Ghosh, Sucharita, and Steven Yamarik. 2004. "Are regional trading arrangements trade creating?: An application of extreme bounds analysis". *Journal of International Economics* 63 (2): 369–395.

Greenwood, Jeremy, Ananth Seshadri, and Guillaume Vandenbroucke. 2005. "The baby boom and baby bust". *American Economic Review* 95 (1): 183–207.

Hartigan, John A, and Manchek A Wong. 1979. "Algorithm AS 136: A k-means clustering algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1): 100–108.

Head, Keith, and Thierry Mayer. 2014. "Gravity equations: Workhorse, toolkit, and cookbook". In *Handbook of international economics*, 4:131–195. Elsevier.

Horn, Henrik, Petros C Mavroidis, and André Sapir. 2010. "Beyond the WTO? An anatomy of EU and US preferential trade agreements". *The World Economy* 33 (11): 1565–1588.

Hussain, Athar. 2002. "Demographic transition in China and its implications". *World development* 30 (10): 1823–1834.

Hutter, I, HB Hilderink, FJ Willekens, and LW Niessen. 1996. "Fertility change in India."

Jones, Charles I. 2001. "Was an industrial revolution inevitable? Economic growth over the very long run". *Advances in macroeconomics* 1 (2).

Kalemli-Ozcan, Sebnem. 2003. "A stochastic model of mortality, fertility, and human capital investment". *Journal of Development Economics* 70 (1): 103–118.

— . 2002. "Does the mortality decline promote economic growth?" *Journal of Economic Growth* 7 (4): 411–439.

Kohl, Tristan, Steven Brakman, and Harry Garretsen. 2016. "Do trade agreements stimulate international trade differently? Evidence from 296 trade agreements". *The World Economy* 39 (1): 97–131.

League Of Nations. *Statistical yearbook: various years.* `http://digital.library.northwestern.edu/league/`.

Lee, Jong-Wha, and Phillip Swagel. 1997. "Trade barriers and trade flows across countries and industries". *Review of Economics and Statistics* 79 (3): 372–382.

McLaughlin, Patrick A, and Oliver Sherouse. 2016. "QuantGovfffdfffdfffdA Policy Analytics Platform". *QuantGov, October* 31.

Mitchell, BR. 1988. "International Historical Statistics: The Americas and Australasia 1750-1988, Detroit, United States: Gale Research Company, 1983., International Historical Statistics". *Europe.*

Mitchell, Brian R. 2003. "International historical statistics: Africa, Asia & Oceania, 1750–2000 . Houndmills, Basingstokes". *Hampshire, New York: Palgrave Macmillan.*

Murphy, Kevin M, Curtis Simon, and Robert Tamura. 2008. "Fertility decline, baby boom, and economic growth". *Journal of Human Capital* 2 (3): 262–302.

OECD. *OECD Better Life Index.* `http://www.oecdbetterlifeindex.org/`.

Potter, Joseph E, Carl P Schmertmann, Renato M Assunção, and Suzana M Cavenaghi. 2010. "Mapping the timing, pace, and scale of the fertility transition in Brazil". *Population and development review* 36 (2): 283–307.

Rele, Jawahar Raghunath. 1988. "70 years of fertility change in Korea: new estimates from 1916 to 1985." *Asia-Pacific population journal* 3 (2): 29–54.

Rose, Andrew K. 2004. "Do we really know that the WTO increases trade?" *American Economic Review* 94 (1): 98–114.

Rosen, Howard. 2004. "Free trade agreements as foreign policy tools: The US-Israel and US-Jordan FTAs". *Free trade agreements: US strategies and priorities*: 51–77.

Salton, Gerard, and Michael J McGill. 1986. "Introduction to modern information retrieval".

Sardon, Jean-Paul. 1991. "Generation replacement in Europe since 1900". *Population an English Selection*: 15–32.

Statistics, National Center for Education. 2010. "Digest of education statistics: 2009". *Tables* 101:186.

Tamura, Robert. 2006. "Human capital and economic development". *Journal of Development Economics* 79 (1): 26–72.

Tamura, Robert, Gerald P Dwyer, John Devereux, and Scott Baier. 2016. "Economic growth in the long run".

Tamura, Robert, Jerry Dwyer, John Devereux, and Scott Baier. 2019. "Economic growth in the long run". *Journal of Development Economics* 137:1–35.

Tamura, Robert, and Curtis Simon. 2017. "Secular fertility declines, baby booms, and economic growth: International evidence". *Macroeconomic Dynamics* 21 (7): 1601–1672.

Tamura, Robert, Curtis Simon, and Kevin M Murphy. 2016. "Black and white fertility, differential baby booms: The value of equal education opportunity". *Journal of Demographic Economics* 82 (1): 27–109.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the number of clusters in a data set via the gap statistic". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2): 411–423.

Tinbergen, Jan. 1962. "Shaping the World Economy The Twentieth Century Fund". *New York*.

Trefler, Daniel. 1993. "Trade liberalization and the theory of endogenous protection: an econometric study of US import policy". *Journal of political Economy* 101 (1): 138–160.

Turner, Chad, Robert Tamura, and Sean E Mulholland. 2013. "How important are human capital, physical capital and total factor productivity for determining state economic growth in the United States, 1840–2000?" *Journal of Economic Growth* 18 (4): 319–371.

Turner, Chad, Robert Tamura, Sean E Mulholland, and Scott Baier. 2007. "Education and Income of the States of the United States: 1840–2000". *Journal of Economic Growth* 12 (2): 101.

Al-Ubaydli, Omar, and Patrick A McLaughlin. 2017. "RegData: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012". *Regulation & Governance* 11 (1): 109–123.

United Nations. *"Demographic Yearbook: various years"*. `https://data.worldbank.org/indicator/`.

United Nations Development Programme. *Human Development Report: various years.* `http://digital.library.northwestern.edu/league/`.

United Nations Population Division. 2015. *World population prospects: The 2015 revision.* Nations, United.

World Bank Group. *World development indicators: various years.* `https://data.worldbank.org/indicator/`.